

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New). Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

8

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2006, 19:01:43 ; Search time 6169 Seconds
(without alignments)
17807.760 Million cell updates/sec

Title: US-10-723-807-1_COPY_49_2396

Perfect score: 2348

Sequence: 1 cctcctctctcttgagcg.....cgacgcgttcgagtcacactg 2348

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:*

1: gb_est1:*

2: gb_est2:*

3: gb_est3:*

4: gb_hic:*

5: gb_est4:*

6: gb_est5:*

7: gb_est6:*

8: gb_est7:*

9: gb_gss1:*

10: gb_gss2:*

11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	303	12.9	878	6	CF818626 EST696008
2	266	11.3	885	6	CF822062 EST699444
3	239.6	10.2	947	6	CF826270 EST703652
4	204.2	8.7	630	8	DR625376 EST101550
5	154.2	6.6	527	10	CF480293 tigr-gss-
6	146.2	6.2	826	7	CO546284 LyEST239
7	146	6.2	784	9	BZ256150 CH230-428
8	143.8	6.1	620	10	CE793096 tigr-gss-
9	137.6	5.9	590	11	CNS03M2Z Tetraodon
10	137.2	5.8	789	9	BZ115869 CH230-255
11	131	5.6	805	11	CNS04RM2 Tetraodon
12	128.6	5.5	774	10	CNS02APC Tetraodon
13	127.2	5.4	572	11	CNS04J3SN Tetraodon
14	127.2	5.4	723	9	AZ627893 1M0469118
15	125.2	5.3	729	7	CV528119 CS GIL 16
16	123.2	5.2	762	10	AG498356 Mus muscu
17	122.4	5.2	646	7	CO543192 LyEST1058
18	120	5.1	633	7	CV070818 CS gil_34
19	118.6	5.1	581	6	CF569196 EST057 Su
20	118.2	5.0	626	9	BZ243402 CH230-419
21	117	5.0	720	11	CR200716 Reverse 8
22	117	5.0	812	11	CR072822 Reverse 8

ALIGNMENTS

RESULT 1

CF818626

LOCUS

DEFINITION

CF818626 EST696008 Coccidioides posadasii saprobic phase cDNA library, greater than 4kb Coccidioides posadasii cDNA clone CIBB578 3' end, mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CF818626 878 bp mRNA linear EST 01-APR-2004
EST696008 Coccidioides posadasii saprobic phase cDNA library,
greater than 4kb Coccidioides posadasii cDNA clone CIBB578 3' end,
mRNA sequence.

CF818626 GI:45924504

EST.

Coccidioides posadasii

Coccidioides posadasii

Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

Onygenales; mitosporic Onygenales; Coccidioides.

1 (bases 1 to 878)

Gardner, M.J. and Cole, G.T.

Analysis of gene expression in Coccidioides posadasii mycelia and

spherules via expressed sequence tags

Unpublished (2003)

Other ESTs: EST696009

Contact: Gardner MJ

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301 838 3519

Fax: 301 838 0208

Email: gardner@tigr.org

Location/Qualifiers

1..878

/organism="Coccidioides posadasii"

/mol_type="mRNA"

/strain="C735"

/db_xref="taxon:199306"

/clone="CIBB578"

/dev_stage="saprobic phase (mycelia)"

/lab_host="E. coli DH10B, T1 phage resistant"

/clone_lib="Coccidioides posadasii saprobic phase cDNA

library, greater than 4kb"

/note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;

Coccidioides posadasii saprobic phase cDNA library, size

fractionated cDNA > 4 kb"

ORIGIN

Query Match 12.9%; Score 303; DB 6; Length 878;
Best Local Similarity 60.2%; Pred. No. 5.2e-67;
Matches 523; Conservative 0; Mismatches 340; Indels 6; Gaps 1;


```
QY 2261 CAGACGAATCCGACGTTGAGTCTGCTTAACATCGCTAGCAAGGAGGTGAAGCAG 2320
Db 595 CAACCAACCCCTACTGTCTCAGTTCCGCTTCTGAATGTGCAATCGAATCGAAGACAAAAGAAAT 654
QY 2321 GCGCGATCGACGCGTTCGAGTCAACTG 2348
Db 655 GTGACACTTGATGCTTTTGAGCGCGATG 682

RESULT 3
LOCUS CF826270 947 bp mRNA linear EST 01-APR-2004
DEFINITION EST703652 Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb Coccidioides posadasii cDNA clone CIDB176 5' end, mRNA
ACCESSION CF826270
VERSION CF826270.1 GI:459332327
KEYWORDS Coccidioides posadasii
SOURCE Coccidioides posadasii
ORGANISM Coccidioides posadasii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE 1 (bases 1 to 947)
AUTHORS Gardner,M.J. and Cole,G.T.
TITLE Analysis of gene expression in Coccidioides posadasii mycelia and spherules via expressed sequence tags
JOURNAL Unpublished (2003)
COMMENT Other ESTs: EST703651
Contact: Gardner MJ
7912 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org
Seq primer: M13 Reverse.

FEATURES
    source
        1..947
            /organism="Coccidioides posadasii"
            /mol_type="mRNA"
            /strain="C735"
            /db_xref="taxon:199306"
            /clone="CIDB176"
            /dev_stage="saprobic phase (mycelia)"
            /lab_host="E. coli DH10B, T1 phage resistant"
            /clone_lib="Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb"
            /note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV; Coccidioides posadasii saprobic phase cDNA library, size fractionated cDNA 2 to 4 kb"

ORIGIN
    Query Match 10.2%; Score 239.6; DB 6; Length 947;
    Best Local Similarity 65.5%; Pred. No. 1.5e-50;
    Matches 350; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

QY 1815 CGGTGCTGTGGGCTCAACCAACACCAAGCAGTATCGGCATTCGTTTTCGGATATT 1874
Db 3 CGGTTCTGTGGGCAACAAATACAAAAACAGTATCGGCATTCCTTTCTTGGGATATC 62

QY 1875 AGCTCCAGATGTGGAGTCACTCAAGTCCGTCCTGATCCCGATCAGGAAGGTGATA 1934
Db 63 ACATTCTGATTTTGAATCTGGCTCAATTGACTCCCTTGACATGACCAACGGGATA 122

QY 1935 TTCAATATGCCAATCGAGCCCGTGGCAATACCATCGCTTTTGTTCGGGAATGACC 1994
Db 123 TCAGTACGCTGCTGGAGTCCCAAGGAAATGTTATCGCTATGTCCGCAATAACAAT 182

QY 1995 TTTATCTGTGGGATAATGGTACCGTTACTCGCATTAATGATGTGTGCCCCGACATGT 2054
Db 183 TATACCTCTGGAAGACGGAGAGGTCACTCAATACGGAAGATGGAGGCCAAATACTT 242

2055 TCACGGCGTGGCGGACTGGATCTATGAGAGGAGATCCTCGGCGATCGTACGCGTTGT 2114
```

```
Db 243 TCAACGCGTGTTCAGACTGGGTATATGAAGAGGAAATTTTCGGTACTCCGTTTGGTTGT 302
QY 2115 GGTTCGCGCAGATGGTGAATATCTGGCTTACTTGGAGCTTCAATGAGACTGGGGTTCCGA 2174
Db 303 GGTTCACCGACGGCGAATATGTGCTTACATGAGGACCGACGAATCGCGTCCCAA 362
QY 2175 CTTACACCGTTTCAGTATTATATGATATGATAACCAAGAGATCGTCCGCGCGTATCCATGGAGC 2234
Db 363 CATACACTATCCCATCTACATGATATCAGAGTTTGGCCCTCCGCTATCCAGAGAAT 422
QY 2235 TGAAGATGAATATCCCAAGGTGTGCGACAGCATCCGACCGTGAAGTTCGCTTGA 2294
Db 423 TGGAGCTTCGTTATCTTAAGTGTGCAACCAACCCCTACTGTCCAGTTCGCCCTTCTGA 482
QY 2295 ACATCGCTAGCAAGGAGGTGAAGCAGGCGCGATCGACGCGTTCGAGTCAACTG 2348
Db 483 ATGTCGAATCCGAGNACAAAAGATGTGACACTTGATGCTTTTGAGGGGATG 536

RESULT 4
LOCUS DR625376 630 bp mRNA linear EST 11-JUL-2005
DEFINITION EST1015504 FvI Gibberella moniliformis cDNA clone FVIBG96, mRNA
ACCESSION DR625376
VERSION DR625376.1 GI:70700088
KEYWORDS Gibberella moniliformis
SOURCE Gibberella moniliformis
ORGANISM Gibberella moniliformis
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE 1 (bases 1 to 630)
AUTHORS Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L, Lee,Y., Utterback,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D., Kendra,D.F., Town,C.D. and Whicelaw,C.A.
TITLE Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the funonisin gene cluster
JOURNAL Unpublished (2005)
COMMENT Contact: Brown, D.W.
USDA/ARS/NCAP
USDA
1815 N. University St, Peoria, IL 61604, USA
Tel: 309 681 6230
Fax: 309 681 6689
Email: brown@ncaur.usda.gov
TIGR sequence name: FVIBG96TH
Seq primer: AAT TAA CCC TCA CTA AAG GG.

FEATURES
    source
        1..630
            /organism="Gibberella moniliformis"
            /mol_type="mRNA"
            /strain="m3125"
            /db_xref="taxon:117187"
            /clone="FVIBG96"
            /tissue_type="mycelia"
            /clone_lib="FvI"
            /note="Vector: pBlueScript II SK(+) XR; Site 1: EcoRI; Site 2: XhoI; anamorph: Fusarium verticillioides. Library FvI was prepared from growth on excised maize seedling roots and shoots. The roots and shoots were prepared as described above and then inoculated by dipping briefly in a suspension of 5 x 106 conidia per ml. The inoculated roots and shoots were then incubated on moistened Whatman #1 filter paper at room temperature under sterile conditions for 4 days. Fungal mycelial was collected, frozen in liquid nitrogen, ground to a powder, and then added to Trizol Reagent (Invitrogen, Carlsbad CA) at approximately 1 g mycelia per 10 ml Trizol. The cDNA was directionally ligated into the pBlueScript II SK(+) XR vector (CDNA Synthesis Kit; Stratagene)."
```

ORIGIN

Query Match 8.7%; Score 204.2; DB 8; Length 630;
Best Local Similarity 60.7%; Pred. No. 2e-41;
Matches 373; Conservative 0; Mismatches 233; Indels 9; Gaps 2;

Qy 1646 GTGCAATGGCTCTCGGCGCAGAGGATGGATCTTACGTACGCGCGGGAAGCGGAGT 1705
Db 20 GTAGAGTGGTCTCGCGAGGTGGAATTAATTAATTTCTCAGGACTCTGGCGAT 79
Qy 1706 CTCACATCGAGAAATCGTCAACCAAGAGTCAAGCGGTCATCTCTGGGGAAGATT 1765
Db 80 CTCGTTCTCGAAGACATCGTCAGCGAAGAGACAAATTCGTACAGCTGATAAACTT 139
Qy 1766 CCGACAGGAAGAGCGTTCAATTAATCTGGATCCGACTTGTCTCGTGGCTGTGG 1825
Db 140 CC-----CAAGGATCTGCACGATCTGGATCAGCAGCGCAAGCGGTTCTCATC 193
Qy 1826 GCGTCCAAACACCAAGCAGTATCGGCATTCGTTCTTTCGCGATTAATAGTCCAGGAT 1885
Db 194 GCGTCAAACTACCAAGCAGTACGGTACTCTGCTACTTCGCGAGATTATTTCACTCGAC 253
Qy 1886 GTGAGTCACTCAAGTCCGTGCCCTGATGCCCGATCAGGAAGTGATATTCATATGCC 1945
Db 254 GTTGAGTCTGCGAGAGTGAACCGCTCGTGAAGACAGGTTGGCGATATCCAGTACGCG 313
Qy 1946 CAATGGAGCCCGTGGGCAATACATCGCTTTTGTTCGCGAGATGACCTTATGTCTGG 2005
Db 314 CAGTTTGGCCGAGTGAGGAAGCGGTTGCGTTTGTGCGGGTAAACAATCTCTTCATTCGT 373
Qy 2006 GAT---AATGGTACCGTTACTTCGCAATCTACTGATGATGTTGGCCCGCATGTTCCACGGC 2062
Db 374 GGTGAGACGCGTGATGTTGACAGATTAACGAGTATGAGGGGCGCGATATGTTTCAAGT 433
Qy 2063 GTGCCGAGTGGATCTATGAAGAGAGATCTCGCGCATCGCTACGCGTTGTGTTCTCG 2122
Db 434 GTACCGGATTCGGTGTATGAGAGAGAGATTTACGGTGTGCTCGGCACTTTGGTGTGCG 493
Qy 2123 CCAGATGTTGAATCTCTGGCTTACTTGGCTTCAATGAGACTGGGGTTCGACCTACACC 2182
Db 494 CCTGATGCCAAGTTCTTTCGCTTTCTGAGCTTTTAAACGAAACTGTTGTTGGGACTTTTACT 553
Qy 2183 GTTCAGATTATATGGAATAACCAAGAGATCGCTCGCGGTATCCATGGGAGCTGAAGATA 2242
Db 554 ATTCCTTACTACATGGAATAACAGAGCTTCGCGGACATATCTCGCGAACTGGACCTT 613
Qy 2243 AGGTATCCCAAGGTG 2257
Db 614 CGGTATCCCAAGGTTG 628

RESULT 5
CE480293/c
LOCUS
DEFINITION
tigr-gss-dog-17000364769283 Dog Library Canis familiaris genomic,
genomic survey sequence.
ACCESSION
CE480293
VERSION
CE480293.1 GI:36797074
KEYWORDS
GSS.
SOURCE
Canis familiaris (dog)
ORGANISM
Canis familiaris
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
REFERENCE
1 (bases 1 to 527)
Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,
Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Frazer, C.M. and
Venter, J.C.
The dog genome: survey sequencing and comparative analysis
Science 301 (5641), 1898-1903 (2003)
14512627
PUBMED
COMMENT
Contact: Kirkness EF
The Institute for Genomic Research
Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
Rockville, MD 20850, USA

Tel: 301-838-0200
Fax: 301-838-0208
Email: ekirknes@tigr.org
Class: shotgun.
Location/Qualifiers
1. 527
/organism="Canis familiaris"
/mol_type="genomic DNA"
/strain="Standard Poodle"
/db_xref="taxon:9615"
/clone_lib="Dog Library"
/notes="Site 1: BstXI; Libraries were prepared from
peripheral blood"

ORIGIN
Query Match 6.6%; Score 154.2; DB 10; Length 527;
Best Local Similarity 60.7%; Pred. No. 1.9e-28;
Matches 252; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Qy 401 GCGGCTCGGCCTTCGTCTCGTGGCAGCCAGTAGACGCTCTCCCTGCTGTTGGGCT 460
Db 527 GGTGGTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 468
Qy 461 GTTGTTCGTGGTCTTCTAGGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 520
Db 467 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 408
Qy 521 GCTGGTAGTGGTACCGGTTGGGCGTGGTGGCAGCCAGTAGACGCTCTCCCTGTTGGG 580
Db 407 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 348
Qy 581 CTTGTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 640
Db 347 GGTGACAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 288
Qy 641 GGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 700
Db 287 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 228
Qy 701 GGAAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 760
Db 227 GGT 168
Qy 761 GCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 815
Db 167 GGT 113

RESULT 6
CE480284/c
LOCUS
DEFINITION
LYEST239 Sea lamprey LyEST Petromyzon marinus cdna, mRNA sequence.
ACCESSION
CE480284
VERSION
CE480284.1 GI:51794584
KEYWORDS
EST.
SOURCE
Petromyzon marinus (sea lamprey)
ORGANISM
Petromyzon marinus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
Petromyzontiformes; Petromyzontidae; Petromyzon.
REFERENCE
1 (bases 1 to 826)
Pancer, Z., Mayer, W.B., Klein, J. and Cooper, M.D.
Prototypic T-cell receptor and CD4-like coreceptor expressed in
lymphocytes of the agnathan sea lamprey
Proc. Natl. Acad. Sci. U.S.A. 101 (36), 13273-13278 (2004)
15328402
PUBMED
COMMENT
Contact: Pancer, Zeev
Division of Developmental and Clinical Immunology
The University of Alabama at Birmingham
378 Wallace Tumor Institute, 1530 Third Avenue, South, Birmingham,
AL 35294-3300
Tel: 205-975-5812
Fax: 205-975-7218
Email: zpancer@uab.edu.

FEATURES
source

Location/Qualifiers
1. 826
/organism="Petromyzon marinus"
/mol_type="mRNA"
/db_xref="taxon:7757"
/cell_type="lymphocyte"
/dev_stage="unstimulated larvae"
/clone_lib="Sea lamprey LyEST"
/note="Vector: Lambda ZAP Express; lymphocyte mRNA ESTs from unstimulated larvae. All are from arrayed colonies from a directionally cloned cDNA library in Lambda ZAP Express (Stratagene). All are single pass 5' sequences."

ORIGIN

Query Match 6.2%; Score 146.2; DB 7; Length 826;
Best Local Similarity 63.0%; Pred. No. 2.4e-26;
Matches 226; Conservative 0; Mismatches 133; Indels 0; Gaps 0;
QY 449 GTCCGTGGCCCTGTTGTTGGTGTCTTAGTGGCGGTGGTCTGGCAGTGGTCTGGT 508
DB 472 GCGCGTGAAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 413
QY 509 GCCAAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 568
DB 412 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 353
QY 569 CCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 628
DB 352 GCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 293
QY 629 GTGCGCAAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 688
DB 292 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 233
QY 689 GTTCTGTGTTGGAGAGCTTCTCAAGGTGCTACTGAGGTCTTCTAGGTGGTGGT 748
DB 232 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 173
QY 749 GGTCTGTGATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 807
DB 172 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 114

RESULT 7
B2256150/c

LOCUS CH230-428N11.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
DEFINITION CH230-428N11, genomic survey sequence.

ACCESSION
B2256150

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;
Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 784)
Zhao, S., Shetty, J., Shatsman, S., Tseng, G., Geer, K.,
Riggs, F., de Jong, P., and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment

REFERENCE
AUTHORS

Unpublished (1999)
Other GSSs: CH230-428N11.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.choi.org).

FEATURES
source

Location/Qualifiers
1. 784
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-428N11"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SNHsd/MCW) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 6.2%; Score 146; DB 9; Length 784;
Best Local Similarity 58.5%; Pred. No. 2.7e-26;
Matches 254; Conservative 0; Mismatches 180; Indels 0; Gaps 0;
QY 379 TGGCGGACTTGACGGTGCACAGGGCGGTCTCGGCCTTCGTCTCGTCGAGCCAGTAGA 438
DB 703 TGGTATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 644
QY 439 CGGTCTCCCTGTTGGGCCCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 498
DB 643 TGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 584
QY 499 TGGTCTGTCGCAAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 558
DB 583 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 524
QY 559 AGACCGTCTCCCTGTTGGGCCCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 618
DB 523 TGGTATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 464
QY 619 CAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 678
DB 463 TGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 404
QY 679 AGTGACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 738
DB 403 TGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 344
QY 739 TGGTATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 798
DB 343 TGGTATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 284
QY 799 TGCTGGTGGCAAT 812
DB 283 TGATGGGGACAAT 270

RESULT 8
CE793096/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

CE793096
tigr-gss-dog-17000317481896 Dog Library Canis familiaris genomic,
genomic survey sequence.
CE793096
CE793096.1 GI:37133863
GSS.
Canis familiaris (dog)
Canis familiaris
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Laurasiatheria; Carnivora; Fissipedia; Canidae;
Canis.
1 (bases 1 to 620)
Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K.,

TITLE	Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
JOURNAL	The dog genome: survey sequencing and comparative analysis
PUBLISHED	Science 301 (5641), 1898-1903 (2003)
COMMENT	14512627 Contact: Kirkness EF The Institute for Genomic Research Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA Tel: 301-838-0200 Fax: 301-838-0208 Email: ekirknes@tigr.org Class: shotgun. Location/Qualifiers 1. 620 /organism="Canis familiaris" /mol_type="genomic DNA" /strain="Standard Poodle" /db_xref="taxon:9615" /clone_lib="Dog Library" /note="Site 1: BstXI; Libraries were prepared from peripheral blood"
FEATURES	source

ORIGIN

Query Match	6.1%;	Score 143.8;	DB 10;	Length 620;
Best Local Similarity	60.3%;	Pred. No. 9.4e-26;		
Matches 255;	Conservative	0;	Mismatches 167;	Indels 1;
				Gaps 1;

[illegible]

RESULT 9	CNS03M2Z	590 bp	DNA	linear	GSS 01-SEP-2000				
CNS03M2Z	LOCUS	Tetraodon nigroviridis genome survey sequence T7 end of clone							
DEFINITION	037C13 of library G from Tetraodon nigroviridis, genomic survey sequence.								
ACCESSION	AL250244								
VERSION	AL250244.1 GI:7971256								
KEYWORDS	GSS; genome survey sequence.								
SOURCE	Tetraodon nigroviridis								
ORGANISM	Tetraodon nigroviridis								
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								

REFERENCE	1	Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fizames,C., Wincker,P., Brottier,F., Quetier,F., Saurin,W. and Weissenbach,J
AUTHORS		
TITLE		Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
JOURNAL		Nat. Genet. 25 (2), 235-238 (2000)
PUBMED		10835645
REFERENCE	2	Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.
AUTHORS		
TITLE		Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
JOURNAL		Genome Res. 10 (7), 939-949 (2000)
PUBMED		10899143
REFERENCE	3	(bases 1 to 590)
AUTHORS		Genoscope.
TITLE		Direct Submission
JOURNAL		Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 131 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
		- Web : www.genoscope.cns.fr)
COMMENT		This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/Tetraodon .

FEATURES
source

```

1. 590
   organism="Tetraodon nigroviridis"
   mol_type="genomic DNA"
   db_xref="taxon:99863"
   clone="037C13"
   clone.lib="G"
   /note="Genoscope sequence ID : COB
end : T3"

```

ORIGIN

Query Match	5.9%	Score 137.6;	DB 11;	Length 590;
Best Local Similarity	57.0%	Pred. No. 3.8e-24;		
Matches 251: Conservative	0;	Mismatches 189;	Indels 0;	Gaps 0;

[illegible]

Matches 204; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

Qy 471 GTGTTCTAGTGGCGGTGGTCTGGCAGTGGTCTGTGGCAAGGTGGTGGTGGTAGTG 530
|||||
Db 542 GTGTTTCTAGTTATGTTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 483
|||||
Qy 531 GTACCGTTGGCGTGGCGGACCCAGTAGACGGTCTCCCTGTTGTTGGCGCTGTTGTTG 590
|||||
Db 482 GTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 423
|||||
Qy 591 GTGCTGTCCTAGTGGCGGTGGTCTGGCAGTGGTCTGGTGCCAAAGGTGGTGGTGGTGA 650
|||||
Db 422 ATAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 363
|||||
Qy 651 GTGTTACCCCTAAGCGCGGTGACGGTCCAGTGGACGGTTCCTGTCGTTGGAGAGCTTG 710
|||||
Db 362 GTGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 303
|||||
Qy 711 CTGAAGTGGTCTACTGGAGGTCTTCTAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 770
|||||
Db 302 CTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 243
|||||
Qy 771 CTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 802
|||||
Db 242 CTGATGTTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 211
|||||

RESULT 15

CV528119 729 bp mRNA linear EST 07-OCT-2004
LOCUS CS_GIL_16F05_SP6 Blue crab gill, normalized Callinectes sapidus
DEFINITION cDNA clone CS_GIL_16F05 5', similar to gb|AA560227.1| phenoloxidase
activating factor - Callinectes sapidus. Score = 140 bits (354),
Expect = 2e-32, mRNA sequence.

ACCESSION

VERSION CV528119

KEYWORDS

SOURCE CV528119.1 GI:53912498

ORGANISM

Callinectes sapidus (blue crab)
Callinectes sapidus
Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Brachyura;
Eubrachyura; Portunoidae; Portunidae; Callinectes.

REFERENCE

1 (bases 1 to 729)
Shafer, T.H., Coblenz, F.E. and Towle, D.W.
Expressed sequence tags from normalized cDNA libraries prepared
from gill and hypodermis tissues of the blue crab, Callinectes
sapidus

JOURNAL

COMMENT Unpublished (2004)

Contact: Thomas H. Shafer
Department of Biological Sciences
University of North Carolina Wilmington
601 S. College Rd, Wilmington, NC 28403, USA

Tel: 910-962-7275

Fax: 910-962-4066

Email: shafer@uncw.edu

Plate: 16 row: F column: 05

Seq primer: SP6

High quality sequence stop: 508.

FEATURES

source

1..729
Location/Qualifiers
/organism="Callinectes sapidus"
/mol_type="mRNA"
/db_xref="taxon:6763"
/clone="CS_GIL_16F05"
/tissue_type="Pooled anterior and posterior gills from
crabs acclimated to salinities of 35 and 5 parts per
thousand"
/dev_stage="Adult intermolt"
/clone_lib="Blue crab gill, normalized"
/note="Vector: pCMV Sport 6.1; Total RNA samples were
prepared individually from each tissue, checked for
quality, and then pooled for construction and
normalization of a cDNA library by Invitrogen. Plasmids

were isolated and inserts sequenced from their 5'-ends by
the Blue Crab Molecular Genetics Laboratory at the
University of North Carolina Wilmington. Traces were
trimmed, compared (BLASTx) to NCBI non-redundant protein
database as of 19 July 2004, and processed for submission
to dbEST by trace2dbEST software (Parkinson, Anthony and
Blaxter, unpublished software)."

ORIGIN

Query Match 5.3%; Score 125.2; DB 7; Length 729;
Best Local Similarity 62.7%; Pred. No. 6.6e-21;
Matches 212; Conservative 0; Mismatches 123; Indels 3; Gaps 1;

Qy 470 GGTGTTCTAGGTGGCGGTGGTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 529
Db 338 GGCGCTGGTGGTGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 397
|||||
Qy 530 GGTACCGTTGGCGGTCTGGTGGCAGCCAGTAGACGGTCTCCCTGTTGTTGGCGCTGTTGTT 589
|||||
Db 398 GGTGCTGGTGGTGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 457
|||||
Qy 590 GGTGCTGTTCTAGTGGCGGTGGTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 649
|||||
Db 458 GCGGCTG---GTGGTGGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 514
|||||
Qy 650 AGTGTACCCCTAAGCGCGGTGGTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 709
|||||
Db 515 GGTGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 574
|||||
Qy 710 GCTGMAAGGTGCTACTGGAGGTCTTCTAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 769
Db 575 GGTGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 634
|||||
Qy 770 GCTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 807
Db 635 GGTGGAGCTGGTGGTGGCGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 672
|||||

Search completed: January 9, 2006, 00:22:05

Job time : 6175 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2006, 22:39:17 ; Search time 1258 Seconds
(without alignments)
154334.419 Million cell updates/sec

Title: US-10-723-807-1_COPY_49_2396

Perfect score: 2348

Sequence: 1 cctctctctctctgagc.....cgacgctgtagtcaactg 2348

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA Main:

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq.*
6: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US11_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2348	100.0	4280	7	US-10-723-807-1
2	374	15.9	2298	6	US-10-369-493-36489
3	305.4	13.0	2326	9	US-10-926-188A-34
4	169.2	7.2	1493	6	US-10-029-386-25133
5	161	6.9	1635	3	US-09-864-761-20241
6	161	6.9	1973	3	US-09-864-761-3471
7	161	6.9	2543	6	US-10-029-386-20536
8	158	6.7	1075	3	US-09-864-761-19241
9	158	6.7	1403	3	US-09-864-761-2513
10	153.8	6.6	1083	6	US-10-029-386-20723
11	148.2	6.3	1972	6	US-10-017-161-1981
12	148.2	6.3	1972	6	US-10-292-798-1629
13	147	6.3	446	3	US-09-864-761-20699
14	145	6.2	3297	9	US-10-450-763-27241
15	141.6	6.0	79977	9	US-10-737-082-58
16	141.6	6.0	79977	9	US-10-765-790-58
17	138.6	5.9	135259	7	US-10-240-425-1585
18	135	5.7	155074	5	US-10-026-188-6
19	134.8	5.7	439892	5	US-10-087-192-454
20	134	5.7	1395	9	US-10-450-763-25328
21	133.4	5.7	666	6	US-10-029-386-25948
22	133.4	5.7	1543	9	US-10-450-763-28010
23	133.4	5.7	38918	6	US-10-017-161-2049

24	133.4	5.7	38918	6	US-10-292-798-1695	Sequence 1695, Ap
25	133.4	5.7	122279	9	US-10-756-149-3596	Sequence 3596, Ap
26	133	5.7	1200	9	US-10-450-763-28632	Sequence 28632, A
27	132.4	5.6	1231	6	US-10-017-161-2047	Sequence 2047, Ap
28	132.4	5.6	1231	6	US-10-292-798-1693	Sequence 1693, Ap
29	132.2	5.6	1794	9	US-10-450-763-13991	Sequence 13991, A
30	131.8	5.6	461	6	US-10-029-386-20388	Sequence 20388, A
31	128.8	5.5	652	6	US-10-029-386-22783	Sequence 22783, A
32	123.6	5.3	830	3	US-09-864-761-19531	Sequence 19531, A
33	123.6	5.3	2109	9	US-10-450-763-17281	Sequence 17281, A
34	122.6	5.2	479	3	US-09-864-761-768	Sequence 768, App
35	121.8	5.2	744802	6	US-10-292-798-1369	Sequence 1369, Ap
36	120.2	5.1	1168	6	US-10-017-161-2179	Sequence 2179, Ap
37	119.8	5.1	486	6	US-10-465-217-13	Sequence 13, Appl
38	119.4	5.1	39344	5	US-10-087-192-1462	Sequence 1462, Ap
39	118.6	5.1	1168	6	US-10-292-798-1825	Sequence 1825, Ap
40	117.2	5.0	574	3	US-09-864-761-228	Sequence 228, App
41	117.2	5.0	669	3	US-09-864-761-17051	Sequence 17051, A
42	117	5.0	597	6	US-10-029-386-20241	Sequence 20241, A
43	117	5.0	1390	6	US-10-017-161-1781	Sequence 1781, Ap
44	117	5.0	1390	6	US-10-292-798-1437	Sequence 1437, Ap
45	117	5.0	2730	9	US-10-450-763-23586	Sequence 23586, A

ALIGNMENTS

RESULT 1

US-10-723-807-1
; Sequence 1, Application US/10723807
; Publication No. US20040171104A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Blinkovsky
; APPLICANT: Kimberly Brown
; APPLICANT: Michael W. Rey
; APPLICANT: Alan Klotz
; APPLICANT: Tony Klyon
; TITLE OF INVENTION: Polypeptides Having Dipeptidyl
; TITLE OF INVENTION: Aminopeptidase Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 5254.200-US
; CURRENT APPLICATION NUMBER: US/10/723,807
; CURRENT FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: US/09/079,592
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 08/857,884
; PRIOR FILING DATE: 1997-05-16
; PRIOR APPLICATION NUMBER: 60/062,892
; PRIOR FILING DATE: 1997-10-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 4280
; TYPE: DNA
; ORGANISM: Aspergillus
US-10-723-807-1

Query Match	100.0%	Score 2348;	DB 7;	Length 4280;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 2348;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	CCTCCTCTCTCTCTGAGCGTTACCGGACTCGTCGCGCTGTCTCCAGTGGCAACGGTAC	60	
Db	49	CCTCCTCTCTCTCTGAGCGTTACCGGACTCGTCGCGCTGTCTCCAGTGGCAACGGTAC	108	
QY	61	GTATCCTGAAACGACATGTAAACCGTTGACCTGATGATTAGTGGCCAGCTGGTGGAAAT	120	
Db	109	GTATCCTGAAACGACATGTAAACCGTTGACCTGATGATTAGTGGCCAGCTGGTGGAAAT	168	
QY	121	CATCGACCGCGATCTTCCCGTCCCTGTCCCTGAGCTCCCTACCAAGGGTCTCCCTATTGT	180	
Db	169	CATCGACCGCGATCTTCCCGTCCCTGTCCCTGAGCTCCCTACCAAGGGTCTCCCTATTGT	228	
QY	181	TGACGGATTGACTGGCGGCAATAAGGGTGGCGAGAGCCTGGAGCAAGGTTACTCTCTCG	240	

|||||
Db 229 TGACGGATTGACTGGCGCAATAAGGGTGGCGAAGCCCTGGGAAGCAAGGTTTACTCTCTCG 288
QY
241 TGAAGACCCCTACCGGACAGCCCTGATGGAAGGGCAATGATGGCCGCCACGCGTATCT 300
Db
289 TGAAGACCCCTACCGGACAGCCCTGATGGAAGGGCAATGATGGCCGCCACGCGTATCT 348
QY
301 TACCGGACGTCCTCGGTCAAGGGGGTCTTGACAACCCCTTTGATCTCCCTACTCCAGAGCT 360
Db
349 TACCGGACGTCCTCGGTCAAGGGGGTCTTGACAACCCCTTTGATCTCCCTACTCCAGAGCT 408
QY
361 TCCTCCCGTCAAGCTTCTCGGCGGACTTGACGGTGGCAAGGGCGGTCTCGGCTTGGTCG 420
Db
409 TCCTCCCGTCAAGCTTCTCGGCGGACTTGACGGTGGCAAGGGCGGTCTCGGCTTGGTCG 468
QY
421 TCCTGGCAGCCAGTACAGCGTCTCCCTGCTGCTGGGCCCTGTTGTTGGTGGTCTTAGG 480
Db
469 TCGTGGCAGCCAGTACAGCGTCTCCCTGCTGCTGGGCCCTGTTGTTGGTGGTCTTAGG 528
QY
481 TGGCGGTGGTCTGGCAGTGGTCTGCTGTCGCAAGGGTGGTCTGGTGGTGGTGGTGG 540
Db
529 TGGCGGTGGTCTGGCAGTGGTCTGCTGTCGCAAGGGTGGTCTGGTGGTGGTGGTGG 588
QY
541 GCCTGGTGGCAGCCAGTACAGCGTCTCCCTGCTGCTGGGCCCTGTTGTTGGTGGTGGT 600
Db
589 GCCTGGTGGCAGCCAGTACAGCGTCTCCCTGCTGCTGGGCCCTGTTGTTGGTGGTGGT 648
QY
601 AGTGGCGGTGGTCTGGCAGTGGTCTGCTGTCGCAAGGGTGGTCTGGTGGTGGTGGT 660
Db
649 AGTGGCGGTGGTCTGGCAGTGGTCTGCTGTCGCAAGGGTGGTCTGGTGGTGGTGGT 708
QY
661 TAAGCGCGGTGACCGGTCCAGTGGACGGTGGTCTGCTGGAGAGCTTGTCTGAAGTGC 720
Db
709 TAAGCGCGGTGACCGGTCCAGTGGACGGTGGTCTGCTGGAGAGCTTGTCTGAAGTGC 768
QY
721 TACTGGAGGTCTTCTAGTGGTGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
Db
769 TACTGGAGGTCTTCTAGTGGTGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 828
QY
781 TGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
Db
829 TGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 888
QY
841 GTGAGGTATCCAACTCCATCCCTTGGTGGCCATTCTGTAATAGCAACAAAGGGGTGG 900
Db
889 GTGAGGTATCCAACTCCATCCCTTGGTGGCCATTCTGTAATAGCAACAAAGGGGTGG 948
QY
901 TACTTGTGCGGATGCTATGCTCTGCGATTGAAGCTAGCGATTCTGTATGTACAAATA 960
Db
949 TACTTGTGCGGATGCTATGCTCTGCGATTGAAGCTAGCGATTCTGTATGTACAAATA 1008
QY
961 ATTTTAAGCAGCTTGGTTCATATCTTCTCACTGGTTTTGGGATATTTTTTCACTT 1020
Db
1009 ATTTTAAGCAGCTTGGTTCATATCTTCTCACTGGTTTTGGGATATTTTTTCACTT 1068
QY
1021 ATTGAATCTTGTAGTAGTCCAGCTTCTCATGGTTAGACACGGGATAACCCCCCAATAGCA 1080
Db
1069 ATTGAATCTTGTAGTAGTCCAGCTTCTCATGGTTAGACACGGGATAACCCCCCAATAGCA 1128
QY
1081 TCATCTGAGGTTGATGTTGCAATGGTCAAGTTTGTCTTAATATGTACGAGTCTTG 1140
Db
1129 TCATCTGAGGTTGATGTTGCAATGGTCAAGTTTGTCTTAATATGTACGAGTCTTG 1188
QY
1141 GGTTACACCGCTAGAGCTTTGCCAACATGAAGCTGTAGCTTGTGCCAAGGCTATCAGC 1200
Db
1189 GGTTACACCGCTAGAGCTTTGCCAACATGAAGCTGTAGCTTGTGCCAAGGCTATCAGC 1248
QY
1201 GGTTTTTTTTATGAGAACTTTGGCAGGATAGGAAAAGTTGGTGGTGGTGAAGGAGCTAAT 1260
Db
1249 GGTTTTTTTTATGAGAACTTTGGCAGGATAGGAAAAGTTGGTGGTGGTGAAGGAGCTAAT 1308
QY
1261 GCAGGAGTGGAGTGAATGAAGCGGATTTCTCGCGGGGAAAAGAAAGGACCAAT 1320
|||||

Db 1309 GCAGGAGTGGAGTGAAGTGAATGAAGCGGATTTCTCGGGGAAAAAGAAAGGACCAAT 1368
QY
1321 TTATGGGACTATTTTATTTTAAACGGGAAGTCTTCAATTCGGCTTCGGCAGGCATCCCTTGA 1380
Db
1369 TTATGGGACTATTTTATTTTAAACGGGAAGTCTTCAATTCGGCTTCGGCAGGCATCCCTTGA 1428
QY
1381 TTCGAGCTGAAGTCTCGGGGTTTTTTTCCACCATGAAGGTAGTCAATTCCTCACTGATTAACA 1440
Db
1429 TTCGAGCTGAAGTCTCGGGGTTTTTTTCCACCATGAAGGTAGTCAATTCCTCACTGATTAACA 1488
QY
1441 TTATTTGTTTACATACACTCCATCATTTGAGTCAATTAATAATAACCTCATTAATTCAGTA 1500
Db
1489 TTATTTGTTTACATACACTCCATCATTTGAGTCAATTAATAATAACCTCATTAATTCAGTA 1548
QY
1501 CTCGAAGCTTCTGCTGCTCTGCTGAGTGGTTCAGGCGCTCGATGTGCTCGGAAC 1560
Db
1549 CTCGAAGCTTCTGCTGCTCTGCTGAGTGGTTCAGGCGCTCGATGTGCTCGGAAC 1608
QY
1561 ACACGCGCCACCGGAGAGGAGTGAAGGCTCTCACCTTCAATGAGACCGTGTGTCGAAGCA 1620
Db
1609 ACACGCGCCACCGGAGAGGAGTGAAGGCTCTCACCTTCAATGAGACCGTGTGTCGAAGCA 1668
QY
1621 AGCAATTAACGCGGACCTCTCGCTCGGTGCAATGGCTCTCGGGCGCAGAGGATGATCCTA 1680
Db
1669 AGCAATTAACGCGGACCTCTCGCTCGGTGCAATGGCTCTCGGGCGCAGAGGATGATCCTA 1728
QY
1681 CGTGTACGCGCGGAGAGACGGCAGTCTCACCATCGAGAAACATCGTCAACCAAGGTCACG 1740
Db
1729 CGTGTACGCGCGGAGAGACGGCAGTCTCACCATCGAGAAACATCGTCAACCAAGGTCACG 1788
QY
1741 CAGCTCATCCCTCGCGACAGGATTCGACAGAGGAAGGAGCGTTCAATTTACTTGGATCCA 1800
Db
1789 CAGCTCATCCCTCGCGACAGGATTCGACAGGAGGAAGGAGCGTTCAATTTACTTGGATCCA 1848
QY
1801 TCCGAGCTTCTGCTCGGTGCTGTGGGCGTCAACACCAACCAAGCAGTATCGGCAATCGTT 1860
Db
1849 TCCGAGCTTCTGCTCGGTGCTGTGGGCGTCAACACCAACCAAGCAGTATCGGCAATCGTT 1908
QY
1861 CTTTGGCGATTAATACGTCAGGATGTGGAGTCACTCAAGTCCGTGCGCTGATGCGCGA 1920
Db
1909 CTTTGGCGATTAATACGTCAGGATGTGGAGTCACTCAAGTCCGTGCGCTGATGCGCGA 1968
QY
1921 TCAGGAGGTGATTAATCAATATGCCAATGAGCGCCGTCGCAATACCAATCGCTTTTGT 1980
Db
1969 TCAGGAGGTGATTAATCAATATGCCAATGAGCGCCGTCGCAATACCAATCGCTTTTGT 2028
QY
1981 TCGCGAATGACCTTTATGCTGGGATAATGGTACCGTTACTCGCAATTTACTGATGATGG 2040
Db
2029 TCGCGAATGACCTTTATGCTGGGATAATGGTACCGTTACTCGCAATTTACTGATGATGG 2088
QY
2041 TGGCCCCGACATGTTCCACGGGTCGCGGACTGATCTATGAGAGGAGATTCCTCGCGCA 2100
Db
2089 TGGCCCCGACATGTTCCACGGGTCGCGGACTGATCTATGAGAGGAGATTCCTCGCGCA 2148
QY
2101 TCGTACGCTGTGGTTCTCGCCAGATGGTGAATATCTGGCTTACTTGGCTTCAATGA 2160
Db
2149 TCGTACGCTGTGGTTCTCGCCAGATGGTGAATATCTGGCTTACTTGGCTTCAATGA 2208
QY
2161 GACTGGGGTTCCGACCTTACACCGTTCACTAGTATTAATATGATAACCAAGAGATTCGCTCGGC 2220
Db
2209 GACTGGGGTTCCGACCTTACACCGTTCACTAGTATTAATATGATAACCAAGAGATTCGCTCGGC 2268
QY
2221 GTATCCATGGAGCTGAAGATTAAGGTATCCCAAGGTGTGCGAGACGATTCGACCGTGAC 2280
Db
2269 GTATCCATGGAGCTGAAGATTAAGGTATCCCAAGGTGTGCGAGACGATTCGACCGTGAC 2328
QY
2281 GTTGAGTCTGTTAACTCGCTAGCAGGAGTCAAGCAGCGCGGATTCGACGGTTCGA 2340
Db
2329 GTTGAGTCTGTTAACTCGCTAGCAGGAGTGAAGCAGCGCGGATTCGACGGTTCGA 2388
QY
2341 GTCAACTG 2348
Db
2389 GTCAACTG 2396

1931	GATATTCAATATGCCCCAATGGAGCCCGTGGGCAATAACATCGCTTTTGTTCGGAGAAT	1990
Qy		
430	GACATCCAGTATGCTCAATGGAGCCCCATGAACAACCTCTATCGCCTATGTCGCTGTRAAC	489
Db		
1991	GACCTTTATCTCTGGGTAATAGTGACCGTTACTTCGCATTAATCTGATGATGTGGCCCGCAC	2050
Qy		
490	GACCTGTATATCTGGAACAATGGCAAGACCAAGCGTATTATCCGAAAAATGGCGGCCCGAT	5496
Db		
2051	ATGTTCCACGGCGTCCGGACTGGATCTATGAAGAGGAGATCCTTCGGCGCATCGTAGCGG	2110
Qy		
550	ATCTTCAAATGGTGCCTGACTGGGTATA CGAGGAAGAAATCTTCGGGGACCGGTTGCCT	609
Db		
2111	TGTGGTTCTCGCAGATGGTGAATATCTGGCTTACTTGTGACTTCAATAGACHTGGGGTT	2170
Qy		
610	CTTTGGTTCTCACTGTACGGTGAATACCTTGGTACCCTCGCTTTAACAGACACTGGAGTC	669
Db		
2171	CCGACCTACACCGTTAGTATTATATGGATAA CCAAGAGATCGCTCCGGCGTATCCATGG	2230
Qy		
670	CCGACCTTACACTATTCCTGTTACTACAGAA CAAGCAAAAGATTGGCCCTCGCTACCCAAGG	729
Db		
2231	GAGCTGAAGATTAAGTATCCCAAGTGTGCGCAGACGAATCCGACCGTGAGCTTGAAGTCTG	2290
Qy		
730	GAGCTGGAGATCCGTTACCTTAAGTCTCTCGGNAGAACCCCAACCGTGCAGTTCACCTG	789
Db		
2291	CTTTAAACATCGCTAGCAAGGAGGTGAAGCAGCGCCGATCGACGCGTTTCGAGTCAA	2345
Qy		
790	TTAAACATTCGCTTCATCCCAAGGAGCAACATATCCAGTTACTGCTGTTCCCGGAAA	844
Db		

RESULT 4

```

US-10-029-386-25133/c
; Sequence 25133, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25133
; LENGTH: 1493
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001751.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.6
US-10-029-386-25133

```

[illegible]

RESULTS

US-09-864-761-20241/C
 ; Sequence 20241, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharron G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; APPLICANT: Chen, Wensheng
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 ; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 ; FILE REFERENCE: Aemica-X-1
 ; CURRENT APPLICATION NUMBER: US/09/864,761
 ; CURRENT FILING DATE: 2001-05-23
 ; PRIOR APPLICATION NUMBER: US 60/180,312
 ; PRIOR FILING DATE: 2000-02-04
 ; PRIOR APPLICATION NUMBER: US 60/207,456
 ; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: US 09/632,366
 ; PRIOR FILING DATE: 2000-08-03
 ; PRIOR APPLICATION NUMBER: GB 24263.6
 ; PRIOR FILING DATE: 2000-10-04
 ; PRIOR APPLICATION NUMBER: US 60/236,359
 ; PRIOR FILING DATE: 2000-09-27
 ; PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00667
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00664
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00669
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00665
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00668
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00662
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00661

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20241
; LENGTH: 1635
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006547.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
; US-09-864-761-20241

Query Match 6.9%; Score 161; DB 3; Length 1635;
Best Local Similarity 50.6%; Pred. No. 3.6e-35;
Matches 389; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

Qy 38 GCTGTCAGCTGGCAACGGTAGCTATCCTGAAACGACATGAGACCCCTGACCTGATGA 97
Db 1300 GGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1241

Qy 98 TTATGAGCCAGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 157
Db 1240 AGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1181

Qy 158 CTTACCAAGGCTCTCCCTATTGTTGACGGATTGACTGGCGGCAATPAAGGGTGGCGAGAAG 217
Db 1180 GGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1121

Qy 218 CTTGGAACGAAGTTACTCTCTGTAAGAACCCCTACCGGACGGCCCTGATGGCGAAGGC 277
Db 1120 GGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1061

Qy 278 AATGATGCCCGACGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 337
Db 1060 GGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1001

Qy 338 TTGATCTCCCTACTCCAGAGCTTCTCCCGTCAAGCTTCTCCGCGGACCTTGACCGTGGC 397
Db 1000 GATGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 941

Qy 398 AAGCGCGGTCTCGGCTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 457
Db 940 GGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 881

Qy 458 CTTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 517
Db 880 GGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 821

Qy 518 GGTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 577
Db 820 GGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 761

Qy 578 GGGCCCTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 637
Db 760 GGTGATGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 701

```

```

Qy 638 GGTGCTGCTGCTAGTGGTACCCCTAAGCCGCTGACCGTCCAGTGGAGCGGTGTTCTCTGTC 697
Db 700 GGTGCTGCTGCTAGTGGTACCCCTAAGCCGCTGACCGTCCAGTGGAGCGGTGTTCTCTGTC 641

Qy 698 GTTGGAGAGCTTGGTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 757
Db 640 GGTGCTGCTGCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 581

Qy 758 GATGCTGCTGCTGCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 806
Db 580 GGTGCTGCTGCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 532

RESULT 6
US-09-864-761-3471/c
; Sequence 3471, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aesomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3471
; LENGTH: 1973
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006547.9
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 11

```

```
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 11
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
US-09-864-761-3471
```

```
Query Match          6.9%; Score 161; DB 3; Length 1973;
Best Local Similarity 50.6%; Pred. No. 4.1e-35;
Matches 389; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

QY 38 GCTGCTCCAGCTGGCAACGGTACGTATCCTGGAACGCAAAATGTAAGACGCTTGACTGATGA 97
DB 1454 GGTGATGGTGTAGTGATGGTGGTTGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGT 1395

QY 98 TTAGTAGCCAGCTGTGGAAATCATCGACGGGATCTCCCGTCCCTGTCCTCGGACTC 157
DB 1394 AGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1335

QY 158 CCTACCAAGGCTCTCCCTATTGTTGACGGATTGACTGGCGCAATAGAGGTGGCGAGAAG 217
DB 1334 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1275

QY 218 CCTGGAAGCAAGGTTACTCTCTGGAAGACCCCTACCGCGACGCGCCCTGATGGCAAGGC 277
DB 1274 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1215

QY 278 AATGATGCCGCCAGCGTGTATCTTACCCGACGTCCTCCCGTCAAGGGGTCTTGACAAACCT 337
DB 1214 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1155

QY 338 TTGATCTCCCTACTCAGAGCTCTCCCGTCAAGCTTCTGGCGACTTGACGGTGGC 397
DB 1154 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1095

QY 398 AAGGCGCGTCTCGGCCCTTCTGTCGTGGAGCCAGTACAGCGTCTCCCTGTCGTTGGG 457
DB 1094 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1035

QY 458 CTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 517
DB 1034 GGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 975

QY 518 GGTGCTGAGTGGTACCGTTGGCGGTCGTGGCAGCCAGTACAGCGTCTCCCTGTTGTT 577
DB 974 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 915

QY 578 GGGCCTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 637
DB 914 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 855

QY 638 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 697
DB 854 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 795

QY 698 GTTGAGAGCTTGTGAGGTGCTACTGAGGCTCTTCTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 757
DB 794 GGTGGTGGCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 735

QY 758 GATGCTGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 806
DB 734 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 686
```

```
RESULT 7
US-10-029-386-20536/c
; Sequence 20536, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
```

```
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20536
; LENGTH: 2543
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007663.28
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.3
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.3
US-10-029-386-20536
```

```
Query Match          6.9%; Score 161; DB 6; Length 2543;
Best Local Similarity 50.8%; Pred. No. 4.8e-35;
Matches 389; Conservative 0; Mismatches 380; Indels 0; Gaps 0;

QY 38 GCTGCTCCAGCTGGCAACGGTACGTATCCTGGAACGCAAAATGTAAGACGCTTGACTGATGA 97
DB 1283 GGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1224

QY 98 TTAGTAGCCAGCTGTGGAAATCATCGACGGGATCTTCCCGTCCCTGTCCTCGGACTC 157
DB 1223 AGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1164

QY 158 CCTACCAAGGCTCTCCCTATTGTTGACGGATTGACTGGCGCAATAGAGGTGGCGAGAAG 217
DB 1163 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1104

QY 218 CCTGGAAGCAAGGTTACTCTCTGGAAGACCCCTACCGCGACGCGCCCTGATGGCAAGGC 277
DB 1103 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1044

QY 278 AATGATGGCCCCAGCGTGTATCTTACCGGACGTCCTCCCGTCAAGGGGTCTTGACAAACCT 337
DB 1043 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 984

QY 338 TTGATCTCCCTACTCCAGAGCTTCTCCCGTCAAGCTTCTTGGCGACTTGACGGTGGC 397
DB 983 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 924

QY 398 AAGGCGCGTCTCGGCCCTTCTGTCGTGGAGCCAGTACAGCGTCTCCCTGTCGTTGGG 457
DB 923 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 864

QY 458 CTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 517
DB 863 GGTAGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 804

QY 518 GGTGCTGAGTGGTACCGTTGGCGGTCGTGGCGAGCCAGTACAGCGTCTCCCTGTTGTT 577
DB 803 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 744

QY 578 GGGCCTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 637
DB 743 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 684

QY 638 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 697
DB 683 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 624

QY 698 GTTGAGAGCTTGTGAGAGGCTTCTACTGAGGCTTCTTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 757
```


;; PRIOR APPLICATION NUMBER: US 09/632,366
;; PRIOR FILING DATE: 2000-08-03
;; PRIOR APPLICATION NUMBER: GB 24263.6
;; PRIOR FILING DATE: 2000-10-04
;; PRIOR APPLICATION NUMBER: US 60/236,359
;; PRIOR FILING DATE: 2000-09-27
;; PRIOR APPLICATION NUMBER: PCT/US01/00666
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00667
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00664
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00669
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00665
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00668
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00663
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00662
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00661
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: PCT/US01/00670
;; PRIOR FILING DATE: 2001-01-30
;; PRIOR APPLICATION NUMBER: US 60/234,687
;; PRIOR FILING DATE: 2000-09-21
;; PRIOR APPLICATION NUMBER: US 09/608,408
;; PRIOR FILING DATE: 2000-06-30
;; PRIOR APPLICATION NUMBER: US 09/774,203
;; PRIOR FILING DATE: 2001-01-29
;; NUMBER OF SEQ ID NOS: 49117
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 2513
;; LENGTH: 1403
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL078472.1
;; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 27
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 19
;; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 34
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 28
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 43
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
;; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 25
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 18
;; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 21
US-09-864-761-2513

Query Match 6.7%; Score 158; DB 3; Length 1403;
Best Local Similarity 53.2%; Pred. No. 2.5e-34;
Matches 335; Conservative 0; Mismatches 295; Indels 0; Gaps 0;

Qy 178 TGTTCACGGATTGACTGGCGGCAATAAGGGTGGCGAGAACCTCGAAGCAAGGTTACTCC 237
Db 674 TGGTGTGATGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 733
Qy 238 TCGTGAAGACCTTACCGGCACGCCCTGATGGCAGGGCAATGATGCCGCCGACGGTGA 297
Db 734 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 793
Qy 298 TCTTACCGAGCTCCCGTCAAGGGGCTTGACAAACCTTTTCGATCTCCCTACTCCAGA 357
Db 794 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 853
Qy 358 CTTCTCTCCCTCAAGCTTCTTCGCGGACTTGAACGGTGGCAAGGCGCGTCTCCGCCCTTCG 417
Db 854 TCGTGGTTTTAGTGTGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 913
Qy 418 TCGTGGTGGACCCAGTAGAGGCTCTCCCTGTCTGGGGCTGTTGTTGGTGGTGTCT 477

Db 914 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 973
Qy 478 AGGTGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 537
Db 974 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1033
Qy 538 TGGGCGTGGGCGAGCCAGTAGACGGTCTCCCTGTTGTTGGGCCCTGTTGTTGGTGGTGG 597
Db 1034 TGATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1093
Qy 598 CCTAGGTGGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 657
Db 1094 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1153
Qy 658 CCCTAAGCGCGGTGACGGTCCAGTGACGGTGTTCCTGTCGTTGGAGAGCTTGTCTGAAGG 717
Db 1154 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1213
Qy 718 TGCTACTGGAGGCTTCTAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 777
Db 1214 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1273
Qy 778 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 807
Db 1274 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1303

RESULT 10
US-10-029-386-20723/c
;; Sequence 20723, Application US/10029386
;; Publication No. US20030194704A1
;; GENERAL INFORMATION:
;; APPLICANT: Penn, Sharron G.
;; APPLICANT: Hanzel, David K.
;; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
;; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
;; FILE REFERENCE: AEOMICA-X-2
;; CURRENT APPLICATION NUMBER: US/10/029,386
;; CURRENT FILING DATE: 2001-12-20
;; NUMBER OF SEQ ID NOS: 34288
;; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
;; SEQ ID NO 20723
;; LENGTH: 1083
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; OTHER INFORMATION: MAP TO AL049849.1
;; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.5
;; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
;; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 12
;; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 10
;; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 8.6
;; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 9
;; OTHER INFORMATION: NT HIT: AL163201.2, EVALUE 2.00e-05
;; OTHER INFORMATION: EST_HUMAN HIT: AW794072.1, EVALUE 6.10e+00
US-10-029-386-20723

Query Match 6.6%; Score 153.8; DB 6; Length 1083;
Best Local Similarity 59.9%; Pred. No. 3.5e-33;
Matches 257; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

Qy 379 TGGCGGACTTTCACCGTGGCAAGCGGCTCTCGGCCCTTCGTCGTCGTCGTCGTCGTCGTCG 438
Db 518 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 459
Qy 439 CGGTCTCCCTGTCTGTTGGGCTGTTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 498
Db 458 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 399
Qy 499 TGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 558

[illegible]

RESULT 13
US-09-864-761-20699
; Sequence 20699, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20699
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC007249.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 14
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 13
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 12
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 12
; OTHER INFORMATION: NT HIT: AL161539.2, EVALUAE 3.70e-00
; PS-09-864-761-20699

```

[illegible]

Qy	705	AGCTTGCTGAAGGTGCTACTGGAGGTCTTCTAGGTGGTGATGCTGGTTCTGCTGATGCTG	764
Db	14747	ATGGTGAGGATGGTGATGGTGATGGTGCGATGGTGATGGTGATGGGGATGATGGTG	14688
Qy	765	CTGGTGCTGATGCTGGTGCTGATGCTGGTGCTGGTGCTGGTGCG	808
Db	14687	GTGATGATGATGATGGCGATGATGGTGAGGATGGTGATGGTGAG	14644

Search completed: January 9, 2006, 03:12:36
Job time : 1262 secs

[illegible]

Mon Jan 9 11:57:22 2006

```
; Publication No. US20050266459A1
; GENERAL INFORMATION:
; APPLICANT: POULSEN, TIM S.
; APPLICANT: NIELSEN, KIRSTEN V.
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND NUCLEIC ACID ANALOG PROBES
; FILE REFERENCE: 09138-6000-00000
; CURRENT APPLICATION NUMBER: US/11/121,086
; PRIOR FILING DATE: 2005-05-04
; PRIOR APPLICATION NUMBER: 60/567,570
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 107
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 61
; LENGTH: 169495
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (70072)..(70171)
; OTHER INFORMATION: a, c, g, t, unknown or other
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (139457)..(157244)
; OTHER INFORMATION: a, c, g, t, unknown or other
US-11-121-086-61

Query Match          3.8%; Score 88.8; DB 7; Length 169495;
Best Local Similarity 53.4%; Pred. No. 3.5e-11;
Matches 186; Conservative 0; Mismatches 162; Indels 0; Gaps 0;

QY 461 GTTGTGTTGGTGTCTAGGTGGCGGTGCTGTCGACGTGGTGTGTCGCCAAGGTTGGT 520
DB 67719 GTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 67660

QY 521 GCTGTAGTGTTACCGTTGGGGTGTGTCGACGCCAGTACGAGTGTCTCCCTGTTGGG 580
DB 67659 GTTGTGCGTTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 67600

QY 581 CTTGTTGTTGGTGTCTAGTGGCGGTGTCGACGGTGGCAGGNGTGGTCCCAAGGT 640
DB 67599 GTAGTGGTGGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 67540

QY 641 GGTGCTGGTGTAGTGTACCCCTAAGCGCGTCCAGTCCAGTGGACGGTGTCTCTGCTT 700
DB 67539 ACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 67480

QY 701 GGAGAGCTTGTGAAGTGTCTACTGAGAGTCTTCTAGGTGGTGTGATGCTGTTCTGCTG 760
DB 67479 GACGCTGTTTGTGATGCTGCTGTTGTTGATGATGCTGTTGATGATGCTGTTGTTGATG 67420

QY 761 GCTGCTGGTGTGATGCTGGTGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 808
DB 67419 GCTGTTGTTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 67372

RESULT 10
US-11-136-527-7399/c
; Sequence 7399, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 361
; LENGTH: 600
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-7399

Query Match          3.6%; Score 85.4; DB 7; Length 2479;
Best Local Similarity 54.2%; Pred. No. 3.6e-11;
Matches 173; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 488 GGTGCTGGCAGTGGTGTGGTCCCAAGGTTGGTGTGTTAGTGTACCGTTGGGCTGCT 547
DB 2301 GCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2242

QY 548 GGCAGCCAGTAGACGGTCTCCCTGTTGTTGGGCCCTGTTGTTGGTGTGTTGTTGTTGTTG 607
DB 2241 GGCATTGTGCTGTTGTCAGCAACTGCTGCTCATTGTTGTTGTAAGTGTCTCTGCTGTC 2182

QY 608 GGTGCTGTGGCAGTGGTGTGGTCCCAAGGTTGGTGTGTTAGTGTACCGTTAAGCCG 667
DB 2181 ACTGCTGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2122

QY 668 CGTGACGGTCCAGTGGACGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 727
DB 2121 GCTGCCGCTGGTGCCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2062
```

```
; ORGANISM: Rattus norvegicus
US-11-136-527-7399

Query Match          3.6%; Score 85.4; DB 7; Length 600;
Best Local Similarity 54.2%; Pred. No. 1.9e-11;
Matches 173; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 488 GGTGCTGGCAGTGGTGTGGTCCCAAGGTTGGTGTGTTAGTGTACCGTTGGGCTGCT 547
DB 422 GCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 363

QY 548 GGCAGCCAGTAGACGGTCTCCCTGTTGTTGGGCCCTGTTGTTGGTGTGTTGTTGTTGTTG 607
DB 362 GGCATTGTGCTGTTGTCAGCAACTGCTGCTCATTGTTGTTGTAAGTGTCTCTGCTGTC 303

QY 608 GGTGCTGTGGCAGTGGTGTGGTCCCAAGGTTGGTGTGTTAGTGTACCGTTAAGCCG 667
DB 302 ACTGCTGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 243

QY 668 CGTGACGGTCCAGTGGACGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 727
DB 242 GCTGCCGCTGGTGCCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 183

QY 728 GGTCTTCTAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 787
DB 182 GCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 123

QY 788 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 806
DB 122 GCGGCTGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 104

RESULT 11
US-11-136-527-3303/c
; Sequence 3303, Application US/11136527
; Publication No. US20050287570A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William M
; TITLE OF INVENTION: Probe Arrays For Expression Profiling of Rat Genes
; FILE REFERENCE: 031896-041000 (AM101086)
; CURRENT APPLICATION NUMBER: US/11/136,527
; CURRENT FILING DATE: 2005-05-25
; PRIOR APPLICATION NUMBER: US 60/574,294
; PRIOR FILING DATE: 2005-05-26
; NUMBER OF SEQ ID NOS: 362830
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3303
; LENGTH: 2479
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-11-136-527-3303

Query Match          3.6%; Score 85.4; DB 7; Length 2479;
Best Local Similarity 54.2%; Pred. No. 3.6e-11;
Matches 173; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 488 GGTGCTGGCAGTGGTGTGGTCCCAAGGTTGGTGTGTTAGTGTACCGTTGGGCTGCT 547
DB 2301 GCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2242

QY 548 GGCAGCCAGTAGACGGTCTCCCTGTTGTTGGGCCCTGTTGTTGGTGTGTTGTTGTTGTTG 607
DB 2241 GGCATTGTGCTGTTGTCAGCAACTGCTGCTCATTGTTGTTGTAAGTGTCTCTGCTGTC 2182

QY 608 GGTGCTGTGGCAGTGGTGTGGTCCCAAGGTTGGTGTGTTAGTGTACCGTTAAGCCG 667
DB 2181 ACTGCTGTGTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2122

QY 668 CGTGACGGTCCAGTGGACGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 727
DB 2121 GCTGCCGCTGGTGCCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2062
```


THIS PAGE BLANK (USPTO)

1987	ACACGCGCCCAACGGGAGACGAGTAAGCGTCTCACCTTCAATGAGACCGTAGTCAAGCA	2046
1621	AGCAATTACCGCCGACCTCTCGCTCGGTGCAATGGCTCTCGGGCGCAGAGGATGGAT - CCT	1679
2047	AGCAATTACCGCCGACCTCTCGCTCGGTGCAATGGCTCTCGGGCGCAGAGGATGGATCCCT	2106
1680	ACGTGTACGGCGGGGAAAGACGGCAGTCTCACCATCGAGAACATCGTCAACCAAGAGTCAAC	1739
2107	ACGTGTACGGCGGGGAAAGACGGCAGTCTCACCATCGAGAAACATCGTCAACCAAGAGTCAAC	2166
1740	GCACGCTCATCCCTGCGGGAACAAGATTCCGACAGGGAAGGCGTTCAATTACTGATCC	1799
2167	GCACGCTCAT - CCTGCGGGAACAAGATTCCGACAGGGAAGGCGTTCAATTACTGATCC	2225
1800	ATCCGAGCTTGTGTCGCGTGTGGGCGTCCAAACACACCAAGACAGTATCGGCATTTGCT	1859
2226	ATCCGAGCTTGTGTCGCGTGTGGGCGTCCAAACACCAAGACAGTATCGGCATTTGCT	2285
1860	TCCTTTCGCCGATTATTACGTCACGAGATGTGAGTCACTCAAGTCGCGTCCCTGATGCCCG	1919
2286	TCCTTTCGCCGATTATTACGTCACGAGATGTGAGTCACTCAAGTCGCGTCCCTGATGCCCG	2345
1920	ATCAGGAAGGTGATATTCAATATATGCCCAATGGAGCCCCGTGGGCAATACATCGCTTTTG	1979
2346	ATCAGGAAGGTGATATTCAATATATGCCCAATGGAGCCCCGTGGGCAATACATCGCTTTTG	2405
1980	TTCCGCGAATGACCTTTATGCTCTGGGATTAATGGTACCGTTACTCGCATTTACTGATGATG	2039
2406	TTCCGCGAATGACCTTTATGCTCTGGGATTAATGGTACCGTTACTCGCATTTACTGATGATG	2465
2040	GTGGCCCCGACATGTTCCAGGGCGTCCGACATGGATCTATGAGAGGAGATCCTCGGCG	2099
2466	GTGGCCCCGACATGTTCCAGGGCGTCCGACATGGATCTATGAGAGGAGATCCTCGGCG	2525
2100	ATCGCTACGGTTGTGTTCTCGCCAGATGGTGAAATATCTGGCTTACTTGAGCTTCAATG	2159
2526	ATCGCTACGGTTGTGTTCTCGCCAGATGGTGAAATATCTGGCTTACTTGAGCTTCAATG	2585
2160	AGACTGGGTTCCGACCTACACCGTTTCAGTTATTATATGGATACCAAGAGATCGCTCCCG	2219
2586	AGACTGGGTTCCGACCTACACCGTTTCAGTTATTATATGGATACCAAGAGATCGCTCCCG	2645
2220	CGTATCCATGGGAGCTCAAGATAAGGTATCCCAAGGTGTCCGACAGGAATCCGACCGTGA	2279
2646	CGTATCCATGGGAGCTCAAGATAAGGTATCCCAAGGTGTCCGACAGGAATCCGACCGTGA	2705
2280	CGTTGAGTCTGCTTTAATCGCTAGCAAGGAGGTGAAGCGAGGCGCGATCGACGCGTTGCG	2339
2706	CGTTGAGTCTGCTTTAATCGCTAGCAAGGAGGTGAAGCGAGGCGCGATCGACGCGTTGCG	2765
2340	AGTCAACTG	2348
2766	AGTCAACTG	2774

RESULT 2		
AAAX00013		
ID	AAAX00013	standard; DNA; 2313 BP.
XX		
XX	AAAX00013;	
XX		
XX		
DT	17-MAR-1999	(first entry)
XX		
DE	Aspergillus oryzae	dipeptidyl aminopeptidase encoding DNA #1.
XX		
KW	prolyl dipeptidyl aminopeptidase;	protein hydrolysate; dough;
KW	flavour enhancer;	palatability; mouthfeel; aroma; crust colour; baking;
KW	animal feed additive;	hydrolysis; ss.
XX		
OS	Aspergillus oryzae.	
XX		
FH	Key	Location/Qualifiers

CDS	1. .2313	
FT	/*tag= a	
FT	/note= "no stop codon given"	
PN	WO9851803-A1.	
XX		
PD	19-NOV-1998.	
XX		
PF	12-MAY-1998; 98WO-US009629.	
XX		
PR	16-MAY-1997; 97US-00857884.	
XX		
PR	20-OCT-1997; 97US-0062892P.	
XX		
PA	(NOVO) NOVO NORDISK BIOTECH INC.	
XX		
PI	Blinkovsky A, Brown K, Rey MW, Klotz A, Byun T;	
XX		
DR	WPI; 1999-045232/04.	
DR	P-PSDB; AAN89589.	
XX		
PT	New dipeptidyl aminopeptidase from Aspergillus oryzae - used to produce	
PT	protein hydrolysates enriched in particular amino acids, useful as	
PT	flavour enhancers, e.g. in doughs.	
XX		
PS	Claim 1; Page 53-55; 77pp; English.	
XX		
CC	The present sequence encodes dipeptidyl aminopeptidase (DPAP) from	
CC	Aspergillus oryzae. DPAP acts synergistically with an aminopeptidase (AP)	
CC	to hydrolyse polypeptides, producing protein hydrolysate (PH), useful in	
CC	foods as flavour enhancer, e.g. in baked goods, enriched in: (a) Ala,	
CC	Arg, Asp, Gly and/or Val, or (b), if the substrate has been deamidated,	
CC	in Glu (free and/or peptide bound), in which case products are useful as	
CC	animal feed additives. DPAP can also be used in flavour - improving	
CC	compositions (optionally containing AP) and in dough pre-mixes, also for	
CC	deactivating enzymes and for converting precursors to mature proteins.	
CC	DPAP increases the level of hydrolysis of proteins and thus of flavour	
CC	development, and a mixture with AP may hydrolyse tripeptides that are	
CC	resistant to either enzyme used alone. PH have improved solubility,	
CC	emulsifying and foaming properties, and products containing them have	
CC	better flavour, palatability, mouthfeel, aroma and crust colour	
XX		
SQ	Sequence 2313 BP; 544 A; 689 C; 601 G; 479 T; 0 U; 0 Other;	
Query Match 36.3%; Score 852.8; DB 2; Length 2313;		
Best Local Similarity 99.8%; Pred. No. 2.2e-184;		
Matches 854; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		
QY	1493 ATTCAGTACTCCAGCTTCTGTGCTCTCTGTCAGTGTGTCACAGGCCCTGGATGTGCTT 1552	
DB	1 ATGAAGTACTCCAGCTTCTGTGCTCTCTGTCAGTGTGTCACAGGCCCTGGATGTGCTT 60	
QY	1553 CGGAAACACACGCGCCACCGGAGAGGAGGAGTCTCACCTTCAATGAGACCGTA 1612	
DB	61 CGGAAACACACGCGCCACCGGAGAGGAGGAGTCTCACCTTCAATGAGACCGTA 120	
QY	1613 GTCAAGCAAGCAATTACGCGACCTCTCGCTCGGTGCAATGGCTCTCGGGCGGAGAGAT 1672	
DB	121 GTCAAGCAAGCAATTACGCGACCTCTCGCTCGGTGCAATGGCTCTCGGGCGGAGAGAT 180	
QY	1673 GGATCTTACGTGTCGCGGCGGAGAGCGGAGTCTCACCATCGAGAACATCGTCACCAAC 1732	
DB	181 GGATCTTACGTGTCGCGGCGGAGAGCGGAGTCTCACCATCGAGAACATCGTCACCAAC 240	
QY	1733 GAGTCAAGCAGCTCATCTCCCTGCGGACAAGATTCCGACAGGAAAGGCGTTCAATTAC 1792	
DB	241 GAGTCAAGCAGCTCATCTCCCTGCGGACAAGATTCCGACAGGAAAGGCGTTCAATTAC 300	
QY	1793 TGGATCAATCCGACTTGTGCTCGGTGCTGTGGCGTCCACACACCAAGCAGTATCGG 1852	
DB	301 TGGATCAATCCGACTTGTGCTCGGTGCTGTGGCGTCCACACACCAAGCAGTATCGG 360	
QY	1853 CATTCGTTCTTTGCGGATTATTACGTCCAGGATGTGGAGTCACTCAAGTCCGTGCCCTG 1912	

Db 361 CATTCTGTTCTTCCGATATTACGTCAGGATGTGGAGTCACTCAAGTCGGTCCCGCTG 420
Qy 1913 ATGCCCATCGAAGGTGATATTCAATATATGCCAATGGAGCCCCGGTGGCAATACCATC 1972
Db 421 ATGCCCATCGAAGGTGATATTCAATATATGCCAATGGAGCCCCGGTGGCAATACCATC 480
Qy 1973 GCTTTTCTCGGAGATGACCTTTATGCTCTGGGATAATGTTACCGTTACTTCGCATTACT 2032
Db 481 GCTTTTCTCGGAGATGACCTTTATGCTCTGGGATAATGTTACCGTTACTTCGCATTACT 540
Qy 2033 GATGATGGTGGCCCCGACATGTTTCCACGGCGTGCAGGACTGGATCTATGAAGAGGAGATC 2092
Db 541 GATGATGGTGGCCCCGACATGTTTCCACGGCGTGCAGGACTGGATCTATGAAGAGGAGATC 600
Qy 2093 CTGCGGATGCTACCGGTGTGTGTTCTCGCAGATGGTGAATATCTGGCTTACTTGAGC 2152
Db 601 CTGCGGATGCTACCGGTGTGTGTTCTCGCAGATGGTGAATATCTGGCTTACTTGAGC 660
Qy 2153 TTCAATCAGACTGGGTTCCGACCTACACGTTTCAGTATTATATGGAATAACCAAGAGATC 2212
Db 661 TTCAATCAGACTGGGTTCCGACCTACACGTTTCAGTATTATATGGAATAACCAAGAGATC 720
Qy 2213 GCTCCGGGTATCCATCGGAGCTCAAGATAAGGTATCCCAAGGTGTGCGAGACGAATCCG 2272
Db 721 GCTCCGGGTATCCATCGGAGCTCAAGATAAGGTATCCCAAGGTGTGCGAGACGAATCCG 780
Qy 2273 ACCGTGACGTTGAGTCTGCTTAAATCGCTAGCAAGAGGTGAAGAGCGGCCGATCGAC 2332
Db 781 ACCGTGACGTTGAGTCTGCTTAAATCGCTAGCAAGAGGTGAAGAGCGGCCGATCGAC 840
Qy 2333 GCGTTCGAGTCAACTG 2348
Db 841 GCGTTCGAGTCAACTG 856

RESULT 3

ADS60815
ID ADS60815 standard; cDNA; 2298 BP.

XX AC ADS60815;

XX DT 02-DEC-2004 (first entry)

XX DE Bacterial polynucleotide #12802.

XX KW Recombinant DNA construct; transformed plant; improved plant property;

XX KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;

XX KW pathogen tolerance; pest tolerance; plant disease resistance;

XX KW cell cycle pathway modification; plant growth regulator;

XX KW homologous recombination; seed oil yield; protein yield; carbohydrate;

XX KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

XX KW bacterial polynucleotide; gene; ss.

XX OS Bacteria.

XX FN US2003233675-A1.

XX PD 18-DEC-2003.

XX PF 20-FEB-2003; 2003US-00369493.

XX PR 21-FEB-2002; 2002US-0360039P.

XX PA (CAOY/) CAO Y.

XX PA (HINKLE/) HINKLE G J.

XX PA (SLATER/) SLATER S C.

XX PA (CHEN/) CHEN X.

XX PA (GOLD/) GOLDMAN B S.

XX XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX XX WPI; 2004-061375/06.

XX XX

PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX Claim 1; SEQ ID NO 36489; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polynucleotide used in
CC the scope of the invention. Note: The sequence data for this patent did
CC not form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 2298 BP; 530 A; 611 C; 655 G; 502 T; 0 U; 0 Other;

Qy Query Match 15.9%; Score 374; DB 13; Length 2298;
Db Best Local Similarity 66.1%; Pred. No. 5e-75;
Matches 558; Conservative 0; Mismatches 280; Indels 6; Gaps 1;

Qy 1502 TCCAAGCTTCTGCTCCCTGTCAGTGTGCTCCAGGCCCTCGATGTCTCGGAAACCA 1561
Db 1 TCCGTCTCTATCCGTTCTCTGGTCAGGTCGCCAGGCCGCGGCTCCATGGAGACCG 60

Qy 1562 CACGCGCCACCGGAGAGGCAGTAAGCGTCTCACCTTCAATGAGACCGTAGTCAAGCAA 1621
Db 61 CGCGAGCTAGAGCGCGCGGATCGAAGCGCTTGACATTTCAAGAACTGTGATTGTGCG 120

Qy 1622 GCATTTAGCCGACCTTCTCGTGTGCAATGCTCTCGGGCGCAGAGGATGATCTCTAC 1681
Db 121 GCGTGTGCGCTTCTTCGATCTCGGTGCAATGGATCGGACTGAGAACGACGGGGAATTAT 180

Qy 1682 GTGTACGCGCGGGAAGACGCGCAGTCTCACCATCGAACAATCGTCACCAACGAGTCAAGC 1741
Db 181 GTGTACGAGGAGGAGGATGAAGCATCAAGATTGAGAGCATTTGTCACCAACCGATCGAG 240

Qy 1742 ACGCTCATCTCCCTCGGACAAGATTTCGACAGGGAAGGAGCGTTCAATTACTGGATCCAT 1801
Db 241 ACATCGTTCCAGCGGAGAGATACCTGCG-----CGATGATACAGCTATTGATTTAGC 294

Qy 1802 CCGGACTTGTGTCGGTGTGTTGGGCGTCCAAACACACGAGATATCGGCATTCGTTTC 1861
Db 295 CCGGACCTGTCCGCGGTGCTGTGGGCCACGAAATTATTAAGCAGATACCGCCACTCGTTC 354

Qy 1862 TTTGCCGATTTATACGTCAGGATGTGAGTCACTCAAGTCCGTCCTGATGCCCGAT 1921
Db 355 TTGCCGATTTACTACATCCAGGAGCTCGAGACGTTGGAGACGGTCCGCTTTGGAGGAT 414

Qy 1922 CAGGAAGTGATATTCAATATGCCAATGGAGCCCCCGTGGGCAATACCATCGCTTTTGT 1981
Db 415 ATGTTGGAGATATTCAATATCGGAATGGAGCCCAAGCGCGGATTCATCGCGTTCGTT 474

Qy 1982 CGCGAGAAATGACCTTTATGTCGTGGGATAAATGGTACCGTTACTTCGCATTTACTGATGATG 2041
Db 475 CGCGCAACCAACTTATGGACTTTGTTCTGACGGGAGTGTGTCAGCGCTATCACCAAGACGGT 534

DE	Aspergillus oryzae polynucleotide SEQ ID NO 1999.	DE	Human genome derived single exon probe #25133.
XX		XX	
KW	Aspergillus oryzae; fermentation; fungus; industrial; EST;	KW	Human; probe; ss; gene expression; single exon probe; microarray;
KW	expressed sequence tag; gene; ss.	KW	alternative splicing event; genomic alteration.
XX		XX	
OS	Aspergillus oryzae.	OS	Homo sapiens.
XX		XX	
PN	W0200279476-A1.	PN	US2003194704-A1.
XX		XX	
PD	10-OCT-2002.	PD	16-OCT-2003.
XX		XX	
XX	22-MAR-2002; 2002WO-IB000890.	XX	03-APR-2002; 2002US-00029386.
PF		PF	
XX		XX	
XX	30-MAR-2001; 2001JP-00098371.	PR	03-APR-2002; 2002US-00029386.
XX		XX	
PA	(NAAD-) NAT INST ADVANCED IND SCI & TECHNOLOGY.	PA	(PENN/) PENN S G.
PA	(NARE-) NAT RES INST BREWING.	PA	(RANK/) RANK D R.
PA	(NORQ) NAT FOOD RES INST MIN AGRIC.	PA	(HANZ/) HANZEL D K.
XX		XX	
PI	Machida M, Akita O, Kashiwagi Y, Kitamoto K, Horiuchi H;	PI	Penn SG, Rank DR, Hanzel DK;
PI	Takeuchi M, Kobayashi T, Kitamoto N, Gomi K, Abe K;	XX	
XX		XX	
XX	WPI; 2003-046817/04.	DR	WPI; 2004-119264/12.
XX		XX	
XX	Detection of expression of specific Aspergillus genes for monitoring the	PT	New human genome-derived single exon nucleic acid probes useful for human
PT	fermentation and growth conditions of the fungus, using DNA probes.	PT	gene expression analysis, for identifying or characterizing alternative
PT		PT	splicing events, for assessing genomic alterations or as tools for
XX		PT	surveying tissues.
PS	Claim 1; SEQ ID NO 1999; 48pp + Sequence Listing; Japanese.	XX	
XX		XX	
XX	The invention relates to a polynucleotide having any of 6006 specific	XX	Claim 1; SEQ ID NO 25133; 80pp; English.
CC	sequences (ABZ50888-ABZ56893), which are expressed by a fungus under	CC	The invention relates to a nucleic acid probe for measuring human gene
CC	specific culture conditions including one or more of eutrophic,	CC	expressions, comprising any of the 27,400 fully defined nucleotide
CC	oligotrophic, solid, early germination, alkaline, high temperature, low	CC	sequences in the specification, or their complements or fragments, and
CC	temperature or maltose culture or polynucleotides stringently hybridising	CC	encoding at least 8 amino acids of any of the 688 amino acid sequences
CC	to these sequences. The polynucleotides are useful for monitoring the	CC	fully defined in the specification. The probe is a single exon probe that
CC	progress of fermentation and the growth conditions of a fungus,	CC	hybridises under high stringency conditions to a nucleic acid molecule
CC	especially of Aspergillus oryzae which is widely used in industrial	CC	expressed in human cells or tissues. Also included are a spatially-
CC	fermentation. Also monitoring for fungal contamination. Note: The	CC	addressable set of single exon nucleic acid probes for measuring human
CC	sequence data for this patent did not form part of the printed	CC	gene expression (comprising a plurality of single exon nucleic acid
CC	specification, but was obtained in electronic format directly from WIPO	CC	probes cited above, where each of the plurality of probes is separately
CC	at ftp.wipo.int/pub/published_pct_sequences	CC	and addressably isolatable or amplifiable from the plurality), a single
XX		CC	exon microarray for measuring human gene expression, a method of
XX		CC	measuring human gene expression, a vector comprising the single exon
XX		CC	probe cited above, an ORF-encoded peptide comprising at least 8
XX		CC	contiguous amino acids of any of the above-mentioned amino acid
XX		CC	sequences (optionally with conservative amino acid substitutions), an
XX		CC	isolated antibody that binds specifically to a peptide cited above,
XX		CC	methods of selling and/or licensing single exon probes or microarrays to
XX		CC	a customer desiring to measure gene expression, a method of providing
XX		CC	human gene expression data by subscription, and a computer-readable
XX		CC	storage medium which contains a database having a plurality of records
XX		CC	(each record including data on the expression of a single exon probe
XX		CC	cited above. The probe, methods and apparatus are useful in gene
XX		CC	expression analysis. The probes may be used as tools for surveying
XX		CC	tissues to detect the presence of expressed messages that contain their
XX		CC	specific exon, or in constructing genome-derived single exon microarrays.
XX		CC	In addition, the probes are used in identifying and characterising
XX		CC	alternative splicing events, in detecting and characterising gross
XX		CC	alterations in the genomic locus that includes their exon, in assessing
XX		CC	smaller genomic alterations, in priming the synthesis of nucleic acids,
XX		CC	or in expressing the ORF-encoded peptide. The present sequence is a human
XX		CC	single exon probe of the invention. Note: The sequence data for this
XX		CC	patent did not form part of the printed specification, but was obtained
XX		CC	in electronic format directly from USPTO at
XX		CC	seqdata.uspto.gov/sequence.html?DocID=20030194704
XX		XX	
XX		XX	Sequence 1493 BP; 517 A; 800 C; 17 G; 159 T; 0 U; 0 Other;
XX		XX	
XX	Query Match	XX	Query Match
XX	Best Local Similarity 100.0%; Pred. No. 4.5e-40;	XX	Best Local Similarity 7.2%; Score 169.2; DB 12; Length 1493;
XX	Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	XX	Matches 342; Conservative 0; Mismatches 288; Indels 0; Gaps 0;
Qy	104 GGGCCAGCTGTGGAATCATCGACGGGATCTTCCGTCCTCCTGCTGACTCCCTTACC 163	Qy	178 TGTGTACGGATTGACTGGCGCAATAAGGGTGGCGAGAGCCTGGAAGCAAGTTACTCC 237
Db	272 GGGCCAGCTGTGGAATCATCGACGGGATCTTCCCGTCCCTGCTGACTCCCTTACC 331		
Qy	164 AAGGTCCTCCCTATTGTTCGGATTGACTGGCGCAATAAGGGTGGCGAGCCTTGA 223		
Db	332 AAGGTCCTCCCTATTGTTCGGATTGACTGGCGCAATAAGGGTGGCGAGCCTTGA 391		
Qy	224 AGCAAGGTACTCTCTGTAAGACCCCTACCGCGAGCGCCCTGTATGGCAAGGCAATGAT 283		
Db	392 AGCAAGGTACTCTCTGTAAGACCCCTACCGCGAGCGCCCTGTATGGCAAGGCAATGAT 451		
Qy	284 GGCCCGGACGGTGATCTTACCGAGCGTCCCGGTCGAAGGGG 323		
Db	452 GGCCCGGACGGTGATCTTACCGAGCGTCCCGGTCGAAGGGG 491		
RESULT 6			
ACH91938/c			
ID ACH91938 standard; DNA; 1493 BP.			
XX			
AC ACH91938;			
XX			
XX			
DT 29-JUL-2004 (first entry)			
XX			

[illegible]

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 8, 2006, 20:01:38 ; Search time 298 Seconds
(without alignments)

14005.754 Million cell updates/sec

Title: US-10-723-807-1_COPY_49_2396

Perfect score: 2348

Sequence: 1 cctctctctctctgagc.....cgacgcttcgagcaactg 2348

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/1/ina/1 COMB.seq:*
- 2: /cgn2_6/ptodata/1/ina/5 COMB.seq:*
- 3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
- 4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
- 5: /cgn2_6/ptodata/1/ina/H COMB.seq:*
- 6: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq:*
- 7: /cgn2_6/ptodata/1/ina/PP COMB.seq:*
- 8: /cgn2_6/ptodata/1/ina/RE COMB.seq:*
- 9: /cgn2_6/ptodata/1/ina/backfiles.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2348	100.0	4280	US-09-079-592-1	Sequence 1, Appli
2	2309.8	98.4	5496	US-09-462-284-1	Sequence 1, Appli
3	161.8	6.9	152132	US-09-949-016-13845	Sequence 13845, A
4	161.8	6.9	152145	US-09-949-016-12371	Sequence 12371, A
5	147	6.3	34230	US-09-949-016-12052	Sequence 12052, A
6	147	6.3	128470	US-09-949-016-13765	Sequence 13765, A
7	145	6.2	39154	US-09-949-016-12384	Sequence 12384, A
8	145	6.2	39154	US-09-949-016-12801	Sequence 12801, A
9	145	6.2	39443	US-09-949-016-14326	Sequence 14326, A
10	145	6.2	39443	US-09-949-016-14327	Sequence 14327, A
11	128.2	5.5	522	US-09-949-016-103758	Sequence 103758, A
12	127	5.4	30656	US-09-949-016-14613	Sequence 14613, A
13	121.8	5.2	10304	US-09-627-465B-1	Sequence 1, Appli
14	119.8	5.1	486	US-09-639-207-13	Sequence 13, Appli
15	118.4	5.0	767677	US-09-949-016-12147	Sequence 12147, A
16	118.4	5.0	767677	US-09-949-016-17361	Sequence 17361, A
17	116.4	5.0	114793	US-10-148-806-3	Sequence 3, Appli
18	114.2	4.9	55195	US-09-949-016-15954	Sequence 15954, A
19	112.8	4.8	142783	US-09-949-016-15127	Sequence 15127, A
20	111.6	4.8	57280	US-09-949-016-11796	Sequence 11796, A
21	111.6	4.8	57280	US-09-949-016-12843	Sequence 12843, A
22	111.6	4.8	57280	US-09-949-016-12844	Sequence 12844, A
23	111.6	4.8	57280	US-09-949-016-12846	Sequence 12846, A
24	111.6	4.8	57280	US-09-949-016-13542	Sequence 13542, A

ALIGNMENTS

RESULT 1

US-09-079-592-1
; Sequence 1, Application US/09079592B
; Patent No. 6864092

GENERAL INFORMATION:

; APPLICANT: Alexander Blinkovsky

; APPLICANT: Kimberly Brown

; APPLICANT: Michael W. Rey

; APPLICANT: Alan Klotz

; APPLICANT: Tony Byun

; TITLE OF INVENTION: Polypeptides Having Dipeptidyl

; TITLE OF INVENTION: Aminopeptidase Activity And Nucleic Acids Encoding Same

; FILE REFERENCE: 5254.200-US

; CURRENT APPLICATION NUMBER: US/09/079,592B

; CURRENT FILING DATE: 1998-05-15

; PRIOR APPLICATION NUMBER: 08/857,884

; PRIOR FILING DATE: 1997-05-16

; PRIOR APPLICATION NUMBER: 60/062,892

; PRIOR FILING DATE: 1997-10-20

; NUMBER OF SEQ ID NOS: 13

; SOFTWARE: FASTSEQ for Windows Version 4.0

; SEQ ID NO 1

; LENGTH: 4280

; TYPE: DNA

; ORGANISM: Aspergillus

US-09-079-592-1

Query Match 100.0%; Score 2348; DB 3; Length 4280;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CCTCTCTCTCTCTGAGCGTTACCGGACTCGTCCGCGCTGCTCCAGCTGGCAACGGTAC	60
DB	49	CCTCTCTCTCTCTGAGCGTTACCGGACTCGTCCGCGCTGCTCCAGCTGGCAACGGTAC	108
QY	61	GTATCTCTGAACGACAAATGTAGACGCTTGACTGATGATTAGGCCACGCTGGTGAAT	120
DB	109	GTATCTCTGAACGACAAATGTAGACGCTTGACTGATGATTAGGCCACGCTGGTGAAT	168
QY	121	CATCGACCGGATCTTCCCGTCCCTGCTCCCTGAGCTCCCTACCAAGGGTCTCCCTATTGT	180
DB	169	CATCGACCGGATCTTCCCGTCCCTGCTCCCTGAGCTCCCTACCAAGGGTCTCCCTATTGT	228
QY	181	TCACGGATTGACTGGCGGCAATAAGGGTGGCGGAGAGCTTGAAGCAAGTTACTCTCTCG	240
DB	229	TCACGGATTGACTGGCGGCAATAAGGGTGGCGGAGAGCTTGAAGCAAGTTACTCTCTCG	288

Qy	241	TGAAGACCTTACCGGACGCGCCCTGATGGAAGGCAATGATGGCCCCGACCGGTGATCT	300
Db	289	TGAAGACCTTACCGGACGCGCCCTGATGGCAAGGCAATGATGGCCCCGACCGGTGATCT	348
Qy	301	TACCGGAGTCCCGGTCAAGGGGTCCTTGACAAACCCCTTCGATCTCCCTACTCCAGAGCT	360
Db	349	TACCGGAGTCCCGGTCAAGGGGTCCTTGACAAACCCCTTCGATCTCCCTACTCCAGAGCT	408
Qy	361	TCCTCCCGTCAAGCTTCCTGCGGACCTTGACGCTGGCAAGGGCGTCTCGGCCCTTCGTGCG	420
Db	409	TCCTCCCGTCAAGCTTCCTGCGGACCTTGACGCTGGCAAGGGCGTCTCGGCCCTTCGTGCG	468
Qy	421	TCGTGGCAGCCCACTAGACGCTCTCCCTGCTGTTGGGCTGTTGTTGGTGGTGTCTTAGG	480
Db	469	TCGTGGCAGCCCACTAGACGCTCTCCCTGCTGTTGGGCTGTTGTTGGTGGTGTCTTAGG	528
Qy	481	TGGCGGTGGTCTGGCAGTGGTGTGGTGCCAAAGGTTGGTGTGGTGGTGGTGGTGGTGG	540
Db	529	TGGCGGTGGTCTGGCAGTGGTGTGGTGCCAAAGGTTGGTGTGGTGGTGGTGGTGGTGG	588
Qy	541	GCCTCGTGGCAGCCCACTAGACGCTCTCCCTGTTGGGCTCTCTCTGTTGGTGGTGGTCT	600
Db	589	GCCTCGTGGCAGCCCACTAGACGCTCTCCCTGTTGGGCTCTCTCTGTTGGTGGTGGTCT	648
Qy	601	AGTGGCGGTGGTCTGGCAGTGGTGTGGTGCCAAAGGTTGGTGGTGGTGGTGGTGGTGGT	660
Db	649	AGTGGCGGTGGTCTGGCAGTGGTGTGGTGCCAAAGGTTGGTGGTGGTGGTGGTGGTGGT	708
Qy	661	TAAAGCGGTGACCGGTCACTGGAAGGTTGCTGTCTGGTGGAGAGCTTCTGTAAGGTGC	720
Db	709	TAAAGCGGTGACCGGTCACTGGAAGGTTGCTGTCTGGTGGAGAGCTTCTGTAAGGTGC	768
Qy	721	TACTGGAAGTCTTCTAGTGGTGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	780
Db	769	TACTGGAAGTCTTCTAGTGGTGGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	828
Qy	781	TGCTGATGCTGGTCTGGTCTGGTGGGCAATAGTCTAAACAAGGGCTTTACGGCATCAAT	840
Db	829	TGCTGATGCTGGTCTGGTCTGGTGGGCAATAGTCTAAACAAGGGCTTTACGGCATCAAT	888
Qy	841	GTGAGGTTATCCAAATCCATGCTGGTGGGCAATGCTGTAATAGCAACAAGAGGGGTGG	900
Db	889	GTGAGGTTATCCAAATCCATGCTGGTGGGCAATGCTGTAATAGCAACAAGAGGGGTGG	948
Qy	901	TACTTGGTCGAGTGCATGCTCCTCGATGGAAGCTAGCGATTCCTGTATGTACAAATA	960
Db	949	TACTTGGTCGAGTGCATGCTCCTCGATGGAAGCTAGCGATTCCTGTATGTACAAATA	1008
Qy	961	ATTTTAAACAGCTTGGTTCCATCTGTTCTTCACTGGTTTTTGGATATTTTTTCACTT	1020
Db	1009	ATTTTAAACAGCTTGGTTCCATCTGTTCTTCACTGGTTTTTGGATATTTTTTCACTT	1068
Qy	1021	ATTGAATCTGTAGTAGTCCAGCTTCTCAAGTTAGACAGGGAATAACCCCCCAATAGCA	1080
Db	1069	ATTGAATCTGTAGTAGTCCAGCTTCTCAAGTTAGACAGGGAATAACCCCCCAATAGCA	1128
Qy	1081	TCATCTGAGGTTTTGATGTTGCCAATGGTCAAGTTTTGCTTAATATGATACGAGTCTTG	1140
Db	1129	TCATCTGAGGTTTTGATGTTGCCAATGGTCAAGTTTTGCTTAATATGATACGAGTCTTG	1188
Qy	1141	GGTTACACCGCTAGAAAGCTTTTGCCACCAATGAAGCTGTAGCTTCTGTCACACGGCTATCAGC	1200
Db	1189	GGTTACACCGCTAGAAAGCTTTTGCCACCAATGAAGCTGTAGCTTCTGTCACACGGCTATCAGC	1248
Qy	1201	GGTTTTTTTTATGAATCTTTGGCAGGATAGAAAGTTGGTGGTGGTGGTGGTGGTGGTGGT	1260
Db	1249	GGTTTTTTTTATGAATCTTTGGCAGGATAGAAAGTTGGTGGTGGTGGTGGTGGTGGTGGT	1308
Qy	1261	GCAGGAGTGGAGTGACTGATAAGACGGAATTTCTGGGGGAAAAAGAAAAGGACCAAT	1320
Db	1309	GCAGGAGTGGAGTGACTGATAAGACGGAATTTCTGGGGGAAAAAGAAAAGGACCAAT	1368
Qy	1321	TTATGGGACTATTTATTTTAAACGGGAGTCTTCAATTCGCTTCGACAGCCATCCCTTGA	1380

Qy	1	CCTCCTCTCTCTTGTGACGGCTTACGGGACTCGTCCCGCTGCTCCAGCTGCGCAACGGTAC	60
Db	428	CCTCCTCTCTCTTGTGACGGCTTACGGGACTCGTCCCGCTGCTCCAGCTGCGCAACGGTAC	487
Qy	61	GTATCTCGAACGACAAATGTAAGA CCGTTGAC TGTGATGTAGTAGGCCCCAGCTGGTGGAA	120
Db	488	GTATCTCGAACGACAAATGTAAGA CCGTTGAC TGTGATGTAGTAGGCCCCAGCTGGTGGAA	547
Qy	121	CATCGACCGCGATCTTCCCGTCCCTGTCCCTGCGACTCCCTACCAAGGGTCTCCCTATTGT	180
Db	548	CATCGACCGCGATCTTCCCGTCCCTGTCCCTGCGACTCCCTACCAAGGGTCTCCCTATTGT	607
Qy	181	TGACGGATTGACTGGCGGCANATAGGGTGGGAGAGCCTCGAGACCAAGGTACTCTCCTCG	240
Db	608	TGACGGATTGACTGGCGGCANATAGGGTGGGAGAGCCTCGAGACCAAGGTACTCTCCTCG	667
Qy	241	TGAAGACCTTACCGGCAGCGCCCTGTAGTGCAAGGGCAATGATGGCCCCGACGGTGATCT	300
Db	668	TGAAGACCTTACCGGCAGCGCCCTGTAGTGCAAGGGCAATGATGGCCCCGACGGTGATCT	727
Qy	301	TACCGGAGCTCCCGGTCAAGGGGGTCTTGACAA CCGTTTCGATCTCCCTACTCCAGAGCT	360
Db	728	TACCGGAGCTCCCGGTCAAGGGGGTCTTGACAA CCGTTTCGATCTCCCTACTCCAGAGCT	787
Qy	361	TCCTCCCGTCAAGCTTCTCGGGGACTTGACGGGTGGCAAGGGCGGTCTCGGCCCTCCGCG	420
Db	788	TCCTCCCGTCAAGCTTCTCGGGGACTTGACGGGTGGCAAGGGCGGTCTCGGCCCTCCGCG	847
Qy	421	TCGTGGGACCCAGTAGACGGTCTCCCTGTGCTGGGCTGTGTGTGGTGGTGTCTTAGG	480
Db	848	TCGTGGGACCCAGTAGACGGTCTCCCTGTGCTGGGCTGTGTGTGGTGGTGTCTTAGG	907
Qy	481	TGGCGGTGGTGTGGCAGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG	540
Db	908	TGGCGGTGGTGTGGCAGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG	967
Qy	541	GGTGTGGGACCCAGTAGACGGTCTCCCTGTGCTGGGCTGTGTGTGGTGGTGTGGTGTGGT	600
Db	968	GGTGTGGGACCCAGTAGACGGTCTCCCTGTGCTGGGCTGTGTGTGGTGGTGTGGTGTGGT	102
Qy	601	AGGTGGCGGTGTGTGGCAGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG	660
Db	1028	AGGTGGCGGTGTGTGGCAGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGGTGTGG	108
Qy	661	TAAAGCCGCTGACCGGTCCAGTGGAGCGGTCTTCTGTGCTGGTAGAGCTTGTCTGAAGGTGC	720

; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12384
; LENGTH: 39154
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(39154)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12384

Query Match 6.2%; Score 145; DB 3; Length 39154;
Best Local Similarity 52.6%; Pred. No. 1.8e-25;
Matches 316; Conservative 0; Mismatches 285; Indels 0; Gaps 0;
QY 227 AAGGTTACTCTCGTGAAGACCCCTACCGCAGCGCCCTGATGCGCAAGGGCAATGATGCC 286
Db 16274 AATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16333
QY 287 CCCGACGGTGATCTTACCGGACGCTCCCGGTCAAGGGGCTTTCACAACTTCGATCTC 346
Db 16334 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16393
QY 347 CCTACTCCAGAGCTTCTCCCGTCAAGCTTCTCGGCGGACTTTCACAACTTCGATCTC 406
Db 16394 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16453
QY 407 CTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 466
Db 16454 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16513
QY 467 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 526
Db 16514 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16573
QY 527 AGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 586
Db 16574 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16633
QY 587 GTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 646
Db 16634 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16693
QY 707 CTTGCTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 766
Db 16754 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16813
QY 767 GGTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 826
Db 16814 AGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16873
QY 827 T 827
Db 16874 T 16874

RESULT 8
US-09-949-016-12801
; Sequence 12801, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:

; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12801
; LENGTH: 39154
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(39154)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12801

Query Match 6.2%; Score 145; DB 3; Length 39154;
Best Local Similarity 52.6%; Pred. No. 1.8e-25;
Matches 316; Conservative 0; Mismatches 285; Indels 0; Gaps 0;
QY 227 AAGGTTACTCTCGTGAAGACCCCTACCGCAGCGCCCTGATGCGCAAGGGCAATGATGCC 286
Db 16274 AATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16333
QY 287 CCCGACGGTGATCTTACCGGACGCTCCCGGTCAAGGGGCTTTCACAACTTCGATCTC 346
Db 16334 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16393
QY 347 CCTACTCCAGAGCTTCTCCCGTCAAGCTTCTCGGCGGACTTTCACAACTTCGATCTC 406
Db 16394 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16453
QY 407 CTGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 466
Db 16454 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16513
QY 467 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 526
Db 16514 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16573
QY 527 AGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 586
Db 16574 GATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16633
QY 587 GTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 646
Db 16634 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16693
QY 707 CTTGCTGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 766
Db 16754 GGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16813
QY 767 GGTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 826
Db 16814 AGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 16873
QY 827 T 827
Db 16874 T 16874


```

; LENGTH: 10304
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (267)
; OTHER INFORMATION: a, t, c or g
US-09-627-465B-1

Query Match      5.2%; Score 121.8; DB 3; Length 10304;
Best Local Similarity 57.1%; Pred. No. 6.2e-20;
Matches 222; Conservative 0; Mismatches 167; Indels 0; Gaps 0;

Qy      418  TCGTCGTGGCAGCCAGCTAGACGGTCTCCCTGTCGTTGGGCTGTGTTCGTGGTGTCT 477
      |||
Db      9897  TGGTGATGCGCGATGGAGATGATTATGATGGTGATGATGATGATGATGATGATG 9956

Qy      478  AGGTGGCGGTGGTGCTGGCAGTGGTGTGGTCCAAAGGTGGTGTCTGTGTACGTACCGT 537
      |||
Db      9957  TGGTGATGATGATGGTGACGGTGGTGATGGTGATGGTGATGATGATGATGATGATGATG 10016

Qy      538  TGGCGGTGGTGGCAGCCAGTAGACGGTCTCCCTGTTGTTGGGCCCTGTTGTTGGTGGT 597
      |||
Db      10017  TGGTGATGGTGCTGATGATGATGCTGATGCTGATGCTGATGCTGATGCTGATGATGATG 10076

Qy      598  CCTAGGTGGCGGTGGTGTCTGCAGTGGTGTGGTCCCAAGGTGGTGTCTGTGTAGTGTG 657
      |||
Db      10077  TGACGGTGATGATGGTGATGATGGTGATGGTGATGCTGATGCTGTTGGTGGTGTGATG 10136

Qy      658  CCCTAAGCGCGTCACGGTCCAGTGGACGGTGTTCTCTGCTGGAGAGCTTGCTGGAAGG 717
      |||
Db      10137  TGGTGATGATGATGATGATGATGATGGTGATGATGGTGATGCTGATGATGATGATGATG 10196

Qy      718  TGCTACTGGAGGTCTTCAGGTGGTGATGCTGGTTCTGCTGATGCTGCTGGTGTGATGC 777
      |||
Db      10197  TGATCATGGTGATGATGATGATGGTGATGGTGATGATGATGATGATGATGATGATGATG 10256

Qy      778  TGGTGCTGATGCTGGTGTCTGGTGTGGTG 806
      |||
Db      10257  TGATGGTGATGATGATGATGGTGATGGTG 10285

```

RESULT 14
US-09-639-207-13/c
; Sequence 13, Application US/09639207
; Patent No. 6815575
; GENERAL INFORMATION:
; APPLICANT: Kazemi-Esfarjani, Parsa
; APPLICANT: Benzer, Seymour
; TITLE OF INVENTION: AN ANIMAL MODEL OF POLYGLUTAMINE
; TITLE OF INVENTION: TOXICITY, METHODS OF USE, AND MODULATORS OF POLYGLUTAMINE
; TITLE OF INVENTION: TOXICITY
; FILE REFERENCE: 06618-686001
; CURRENT APPLICATION NUMBER: US/09/639,207
; CURRENT FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: US 60/148,934
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: US 60/148,933
; PRIOR FILING DATE: 1999-08-12
; PRIOR APPLICATION NUMBER: US 60/177,047
; PRIOR FILING DATE: 2000-01-18
; PRIOR APPLICATION NUMBER: US 60/205,720
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 486
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic DNA
US-09-639-207-13

```

Query Match          5.11%; Score 119.8; DB 3; Length 486;
Best Local Similarity 58.7%; Pred. No. 5.2e-20;
Matches 227; Conservative 0; Mismatches 157; Indels 3; Gaps 1;

QY 447 CTGCTGTTGGCCCTGTTGTTGGTGGTCTTCTAGGTGCGGTGCTGTCGGCAGTGGTCTG 506
DB 435 CTGCTGTTGTTGCTGCTGCTGTTGCTGTTGTCAGCTGCTGTTGTTGCTGCTGCTGTTGCTG 376
QY 507 GTGCCAAGGTGCTGCTGTTAGTGGTACCTGTTGGGCGTCTGGCAGCCAGTAGACGGTC 566
DB 375 TTGCTGCTGCTGTTGCTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGTTG 316
QY 567 TCCTGTTGTTGGCCGTTGTTGGTGGTCTCTAGTGGCGGTGGTGGTCTGGCAGTGGTG 626
DB 315 ---CTGTTGCTGCTGCTGCTGTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 259
QY 627 CTGCTGCCAAGGTGCTGCTGTTAGTGGTACCCCTAAGCGCCGTGACGGTCCAGTGGACG 686
DB 258 CTGTTGCTGTTGTTGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGTTG 199
QY 687 GTGTTCTGCTGTTGGAGACCTTGCTCAAGGTGCTACTGAGAGTCTTCTTAGTGGTGAATG 746
DB 198 CTGCTGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 139
QY 747 CTGTTTCTGCTGATGCTGCTGGTGGTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 806
DB 138 TTGCTGCTGCTGTTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 79
QY 807 GGCATAGTCTAACAGGGCTTTTACGG 833
DB 78 CTGTTGCTGCTGCTGGGGGGTTGACGG 52

RESULT 15
US-09-949-016-12147/c
; Sequence 12147, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12147
; LENGTH: 767677
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(767677)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-12147

```

	Query Match	5.0%	Score 118.4;	DB 3;	Length 767677;
	Best Local Similarity	56.3%;	Pred. NO. 2.9e-18;		
	Matches 241;	Conservative	0;	Mismatches 186;	Indels 1; Gaps 1;
QY	379	TGGCGGACTTCACGGTGCACAGGGGGGCTCGGCCCTTCGTCTGCTGGCAGCCCACTAGA	438		
Db	243461	TGGAGTGCTCATGTGTGAGGTGCCCGTGGTGTGGAGGTGCCCTGGTGGTGAAGGTGC	243402		
QY	439	CGGTCCTCGTCTGGTGGCCCTGTGTGGTGGTCTTACGTGSCGGTGGTGTCTGGCAG	498		
Db	243401	CCGTGGTGGTGGAGGTGCTCGTGGTGGTGGACGTGCTCGTGTGGAGGTGCTCTGGTGG	243342		

Qy	499	TGGTCTGGTGC	CAAGGTG	TGCTGTG	TAGTGGT	ACCGTTGGG	CCCTCGTGG	CAGCCCA	GT	558	
Db	243341	AGTTGCTCGT	GGTGGAG	TGCTCGT	GGTGGT	-GAGGTG	TCGTGGT	TGGTGGAG	TT	243283	
Qy	559	AGACG	GTCCCTG	TGTTGG	CCGTGTTGG	TGGTCT	TAGGTGG	CGGTGG	TGCTGG	618	
Db	243282	TCATGGT	GCAGGTG	CTCCTGG	TGGAGATG	CTCGTGG	TGGTGG	TGAGTGT	TGATGG	243223	
Qy	619	CAGTGG	TGCTG	GTGCCA	AGGGTGG	TGCTGG	TAGTGTAC	CCCTAAG	CGCCGTG	ACGGTCC	678
Db	243222	AGGTG	CTCCTGG	TGTGAG	GGTGCTCC	TGGTGG	AGGTGCTCGTGG	TGAGGTG	CTCGTGG	243163	
Qy	679	AGTGG	ACGGTGT	CTGTCGT	TGGAGAG	CTTGCTG	AAAGTGT	CTACTGG	AGGTCTT	CTAGG	738
Db	243162	TGGTGG	AGATGCTCGTGG	TGATGG	AGGTGCTCGTGG	AGGTGCTCGTGG	TGGTGG	AGGTGCTCGTGG	TGGAGG	243103	
Qy	739	TGGTGA	NTGGTTC	GTCTGA	NTGCTGG	TGCTGA	NTGCTGG	TGCTGA	NTGCTGG	TGCTGG	798
Db	243102	TTCTCAT	GGTGCAG	GTGCTCCTGG	TGGAGATG	CTCTGGTGG	TGGAGATG	CTCTGGTGG	TGGAGATG	CTCTGGTGG	243043
Qy	799	TGCTGG	TG	806							
Db	243042	TGGAG	GGT	243035							

Search completed: January 9, 2006, 00:27:13
Job time : 304 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2006, 02:50:35 ; Search time 24 Seconds
(without alignments)
3026.819 Million cell updates/sec

Title: US-10-723-807-2_COPY_17_771

Perfect score: 4033

Sequence: 1 LDVPRKPHAPTGGSKRLTF.....EKNROTQVLHQTWKDLEE 755

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 80:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1162	28.8	793	T41703	dipeptidyl aminope
2	1054	26.1	818	A30107	dipeptidyl aminope
3	981.5	24.3	766	CDH026	dipeptidyl-peptida
4	981	24.3	760	S23752	dipeptidyl-peptida
5	949.5	23.5	792	A39914	dipeptidyl-peptida
6	902.5	22.4	931	A49737	dipeptidyl aminope
7	848.5	21.0	759	I38593	fibroblast activat
8	811.5	20.1	711	S66261	X-Pro dipeptidyl-p
9	779.5	19.3	803	I68600	dipeptidyl aminope
10	779.5	19.3	865	I54331	dipeptidyl aminope
11	773.5	19.2	803	A41793	dipeptidyl aminope
12	747	18.5	738	A87516	dipeptidyl peptida
13	708	17.6	741	JC5142	X-Pro dipeptidyl-p
14	584	14.5	743	T37700	probable dipeptid
15	545	13.5	779	T25173	hypothetical prote
16	545	13.5	799	T25174	hypothetical prote
17	508	12.6	829	T19514	hypothetical prote
18	430.5	10.7	931	T32919	hypothetical prote
19	344.5	8.5	795	F82858	dipeptidyl-peptida
20	243	6.0	657	E70025	probable acylamino
21	210.5	5.2	631	H75007	probable acylamino
22	189.5	4.7	618	D87651	prolyl oligopeptid
23	189.5	4.7	674	B84381	acylaminoacyl-pept
24	187.5	4.6	591	H72474	probable acylamino
25	186	4.6	598	F84199	hypothetical prote
26	186	4.6	642	C71137	hypothetical prote
27	184	4.6	721	T09631	probable acylamino
28	179.5	4.5	709	B82580	alanyl dipeptidyl
29	171	4.2	683	E87495	prolyl oligopeptid

RESULT 1

T41703

dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004

C:Accession: T41703

R:Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, August 1998

A:Reference number: Z22011

A:Accession: T41703

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-793 <MUR>

A:Cross-references: UNIPROT:O14073; UNIPARC:UPI000013AB95; EMBL:AL031180; PIDN:CAA20138

A:Experimental source: strain 972h-; cosmid c2E11 -chimeric

C:Genetics:

A:Gene: SPAC2E11.08

A:Map position: 1

C:Superfamily: dipeptidyl-peptidase IV

Query Match 28.8%; Score 1162; DB 2; Length 793;
Best Local Similarity 36.1%; Pred. No. 1.7e-67;
Matches 272; Conservative 142; Mismatches 294; Indels 46; Gaps 23;

QY	22	ETVVKQAITPTSRSVQWL--SGAEDGSYYVAAEDGSLTIENIVTNESTRILIP-ADKIPTG	78
DB	58	DDIVLQKYKPSYKQVNWIDSQLKD---TFLVKYGLINIQDPYNLKNLFSVSDLYVNG	114
QY	79	--KEAFNYWIHPDLSVLWASNHTKQYRHSFFADYYVQDVESLKSVPML--PDQEGDIQY	134
DB	115	IQLDYDSYISFPDAKYVLYSVNKSQRWRHSSPAQYLYNTET-KDYNMLQGDNEHWTISL	173
QY	135	AQMSPVGNVITAFVRENDLYVWN-GVTYRITDDGGDFMGVDVWLYEEILGDRVALWP	193
DB	174	AEWSPYTGHLQSLFYNNKNDGNVQLTYDGTVDVFNGLTDMWLYEEVLSPPSTIWW	233
QY	194	SPDGEVAYLSFNETGVTYVQVYVNDNOEIAPAYP-----WELKIRYKPVQNTPTVL	248
DB	234	SPDSOKIAFLKUNSEIPTY--HYPLYTAEIDSLPEFDYKNKDMAIKYKPGNPNFSVSL	291
QY	249	SLINIAISKEVKQAPIDAFES-TDLIIGEVAWLTDTHTTVAAKAFNRVQDQKVVAVDTAS	307
DB	292	FVADLNSNASSNFSLWHEPLAEPPVQVNLVWNT--SSVLVQPTNRNSTCITARLLDTEL	349
QY	308	NKATVLSDRDGTGDLNLLSMKYIGPIKPS-----DKDAYIIDISDHSWAHLIYLPVS	362
DB	350	KSHTVTKTECLEBEGWYEVQSQAKWF-PLNLSLWENWSDGYF-DIALDDYNNHLAFIPFN	407
QY	363	GGEPILTQKDMVET-SILSIDOERQLVYVLSQHHSTERHLYSVSYSPFVTPLVDD-T	420
DB	408	GSSPIYLTSGAMVDTPGIHIDGDFGNVFLATLKDSTERHLYVSLDTLEIYGITDNGE	467

ALIGNMENTS


```
Qy 421 VAAVWSAFSANSYGYLYTTCGPDVPPYQELYYTNTSKPLRTITONAKVLQIKDYALPNI 480
Db 468 DEGYSTSFSPFGDFVYLVNHPGDPVWQELASTKDKOYCLSLNETNSRLKQOLSSITLPSV 527
Qy 481 TYFELPLSPGGLTNVMQRLPPGSPDKKYPILETYPYCGRGAQVETKRWQALNFKAYVASD 540
Db 528 EYGLKTF-NDTTFMERRPRNFVNKKYPVLFFAYGPGGQQVAKLFR-VDFQAYLASH 585
Qy 541 SELEYVVTVDNRGTGFKGRKFRSAVTRQLGLLEAEQIYAAQQAANI PMIDADHIGIWG 600
Db 586 PDPEFIVTLDGRGTGFNGAFRYSVSRHLGEMESYDQGQAGKFWADLPFDVDENHVGIWG 645
Qy 601 WSPFGYLTSLVKLEKDSGAFTLGVTAPVSDWRFYDSMYTERYMTKLTSTNEBGETYSVRK 660
Db 646 WSYGYLTSLKLTQD-VFSYGMVAPVTDWRLYDSVYTERYMDLPQYTKBEGYKNSQIHD 704
Qy 661 TDGFKNVGGFLIOHGTGDDNVHQSAAALVDLLMGDGVSPKLSHs-----QWFTSDHGIS 717
Db 705 YEKFKQLK-REFVAHGTGDDNVHFOHSHLM-----DGLNLANCYNDMAVFPDPSAHSIS 758
Qy 718 YHGGGVFLYKQALARKLYQEKNR---QTQVLMHQW 748
Db 759 YHNASLSIYHRLSEWIGDALGRIDPSTGVRQHRW 792

RESULT 2
A30107
dipeptidyl aminopeptidase B (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YHR028C
C;Species: Saccharomyces cerevisiae
C;Date: 07-Jun-1990 #sequence_revision 30-May-1997 #text_change 09-Jul-2004
C;Accession: S46780; A30107
R;Du, Z.
submitted to the EMBL Data Library, June 1994
A;Description: The sequence of S. cerevisiae cosmid 8082.
A;Reference number: S46773
A;Accession: S46780
A;Molecule type: DNA
A;Residues: 1-818 <DUZ>
A;Cross-references: UNIPROT:P18962; UNIPARC:UPI0000031A5F; EMBL:U10399; NID:g500689; PID
R;Roberts, C.J.; Pohlrig, G.; Rochman, J.H.; Stevens, T.H.
J. Cell Biol. 108, 1363-1373, 1989
A;Title: Structure, biosynthesis, and localization of dipeptidyl aminopeptidase B, an in
F;30-45/Domain: transmembrane #status predicted <TM>
A;Reference number: A30107; MUID:89174971; PMID:2647766
A;Accession: A30107
A;Molecule type: DNA
A;Residues: 1-82, 'H', 84-124, 'N', 126-181, 'LRRLET', 189-199, 'N', 201-365, 'DFKRGKERKF', 376-57
A;Cross-references: UNIPARC:UPI0000172A3C; EMBL:X15484
A;Note: the authors translated the codon ACC for residue 572 as Asn
C;Genetics:
A;Gene: SGD:DAP2; STE13; MIPS:YHR028C
A;Cross-references: SGD:S0001070; MIPS:YHR028C
A;Map position: 8R
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein; yeast vacu
F;30-45/Domain: transmembrane #status predicted <TM>
F;63,79,110,139,392,421/Binding site: Carbohydrate (Asn) (covalent) #status predicted

Query Match 26.11%; Score 1054; DB 1; Length 818;
Best Local Similarity 33.11%; Pred. No. 1.9e-60;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEQSKRLTFNETVVKQAITPTTSRVQWLS-----GARDGSYV 48
Db 48 IPHSHNTPDYQEPNSNYNDGKLKVSFS-VVRNNTFPQKYHLEQWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNTSESTLIPADKIPTGKEAFNWIH-----PDLSSVLW 95
Db 107 TFMND-SYVWQSVYDSDSVSLLEGKT-----FIHQNLQVATESITASPDLLRLI 156
Qy 96 ASNHTKQYRHSFADYYVDVSLKSVPLMPDQEQIOYAQWSVPGNTIAFVRENDLYVW 155
Db 157 RTNSVQNRHSHTFGSYFYVD-----KSSSSFEISGNEVALAIWSPNSNDIAYQDNNIYY 212
```

```
Qy 156 ---DNGTIVTITDDGGPDMEHGVDPDKIYEELIGDRVALWFSPDGEVLAYLSFNETGVPT 212
Db 213 SAJSKKTIRAVTNDGSGFLFNGKPDWVYEEVEFEDDKAAWWSPTGDTLAFLEKIDSEVGE 272
Qy 213 YTVQYVMDQEOIAPAPWELKIRYPKVQSQTNPVTLSLLNIASKEVQAIPDAFEST-DL 271
Db 273 FIPIYVQDEK--DIPEMRSIKYPKSGTNPFAELWVYSMKDGTSPHPRI SGNKDGSL 330
Qy 272 IIGEVAWLTDTHTTAAKAFNRVQDQOKVAVDTASNKATVI SDRDGTDCDGLNDLLSMKY 331
Db 331 LITEVTWVGNGN-VLVKTTDRSGDILTFLVIDTIAKTSNVVRNNESSNGGWWEITHNTLF 388
Qy 332 IGPIKSPDKDAY-YIDISDSHGAHLVLPVSGEPI-PLTKGDWEVTS-ILSDQERQL 388
Db 389 IPANETDFRPHNGVVDILPIGGYHNLAYFENSNSHYKTLITEGKWVVGNGLAFDSMNR 448
Qy 389 VYVLSLTOHSTERHLXSVSY-STFAVTPLVDDTVAAVWSAFSANSYGYLYTTCGPDVPPY 447
Db 449 LYFISTRSKSTERHVYIIDLRSNEIIEVTDTSEDGYVDVFSFGRRFGLLTYKGPVY 508
Qy 448 QELYTTNST-----KPLRTITDNAKVLEQIKDYALPNIYFEPLPSPGE---- 491
Db 509 QKIVDFHSRKAECCKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSPRELNLGKDFGKD 568
Qy 492 -TLNMORLPPGPSP--DKKYPILFTYPGGGAQVETKRWQALNFKAYVASDSELYVTW 548
Db 569 ILNYSIELPNDDEFILSDHYPVFFFAYGFPNSQOVVKTF-SVGFNEVVA--SQNLAIIV 625
Qy 549 TVDNRGTFGRKFRSAVTRQLGLEAEQIYAAQQAANI PMIDADHIGIWMGVSFGGYLT 608
Db 626 VVDRGTFGRKQDFRSLVRDLGLYEARDQISAAISYLSLTFFVDPQKLSLFGWSYGGYLT 685
Qy 609 SKVLKDSGA-FTLGVTITAPVSDWRFYDSMYTERYMTKLTSTNEBGETYSVARKTDGFKNV 667
Db 686 LKLTLEKGGRRHFKYGMVA PVTDWRFYDSVYTERYMT PQENFDGYVESSVHNVTALAA 745
Qy 668 EGGFLIOHGTGDDNVHQSAAALVDLLMGDGVSPKLSHsQWFTSDHGISYHGGGVFLYK 727
Db 746 -NRELLMHGTGDDNVHVFQNSLKFLDLDLLNGVENYDVHV--FPDSDHSIRYHNANVI VFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 3
CDHU26
dipeptidyl-peptidase IV (EC 3.4.14.5) - human
N;Alternate names: cell surface glycoprotein CD26; thymocyte-activating molecule (THAM)
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 23-Aug-1996 #text_change 09-Jul-2004
C;Accession: S24313; B42408; A42408; B61136; S59510; I56154; S59857; S15520
R;Misumi, Y.; Hayashi, Y.; Arakawa, F.; Ikehara, Y.
Biochim. Biophys. Acta 1131, 333-336, 1992
A;Title: Molecular cloning and sequence analysis of human dipeptidyl peptidase IV, a ser
A;Reference number: S24313; MUID:92329551; PMID:1352704
A;Accession: S24313
A;Molecule type: mRNA
A;Residues: 1-6, 'I', 8-766 <MIS>
A;Cross-references: UNIPROT:P27487; UNIPARC:UPI000016AE80; EMBL:X60708; NID:g35335; PIDN
R;Darmoul, D.; Lacasa, M.; Baricault, L.; Marguet, D.; Sapin, C.; Trotot, P.; Barbat, A
J. Biol. Chem. 267, 4824-4833, 1992
A;Title: Dipeptidyl peptidase IV (CD 26) gene expression in enterocyte-like colon cancer
IV mRNA levels during cell differentiation.
A;Reference number: A42408; MUID:92165847; PMID:1347043
A;Accession: B42408
A;Molecule type: mRNA
A;Residues: 1-5, 'R', 7-436, 'S', 438-556, 'I', 558-662, 'E', 664-766 <DAR1>
A;Cross-references: UNIPARC:UPI0000052ACB; GB:M80536; NID:g181569; PIDN:AAAS2308.1; PID:
A;Experimental source: intestine
A;Note: this sequence corresponds with the author's translation
A;Accession: A42408
```

A;Molecule type: mRNA
A;Residues: 1-5,'R',7-436,'S',438-556,'I',558-662,'E',664-711,'G',713-766 <DAR2>
A;Cross-references: UNIPARC:UPI0000172A2B; GB:M80536; NID:g181569
A;Note: sequence extracted from NCBI backbone (NCBI:83986; NCBI:83988); this sequence
R;Gorvel, J.P.; Ferrero, A.; Chambraud, L.; Rigal, A.; Bonicel, J.; Maroux, S.
Gastroenterology 101, 618-625, 1991
A;Title: Expression of sucrose-isomaltase and dipeptidylpeptidase IV in human small intestine
A;Reference number: A61136; MUID:91317403; PMID:1677636
A;Accession: B61136
A;Molecule type: protein
A;Residues: 1-15,'X',17-22 <GOR>
A;Cross-references: UNIPARC:UPI0000172A2C
R;Boehm, S.K.; Gum Jr., J.R.; Erickson, R.H.; Hicks, J.W.; Kim, Y.S.
Biochem. J. 311 835-843, 1995
A;Title: Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a
A;Reference number: S59510; MUID:96067599; PMID:7487939
A;Accession: S59510
A;Molecule type: DNA
A;Residues: 1-31 <BOE>
A;Cross-references: UNIPARC:UPI000016B4A6; GB:S79876; NID:g1195574; PIDN:AA35614.1; PID
R;Tanaka, T.; Camerini, D.; Seed, B.; Torimoto, Y.; Dang, N.H.; Kameoka, J.; Dahlberg, H.
J. Immunol. 149, 481-486, 1992
A;Title: Cloning and functional expression of the T cell activation antigen CD26.
A;Reference number: I56154; MUID:92325476; PMID:1352530
A;Accession: I56154
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: mRNA
A;Residues: 1-436,'S',438-766 <TAN>
A;Cross-references: UNIPARC:UPI000004F7BF; GB:M74777; NID:g180082; PIDN:AA51943.1; PID:
R;Abbott, C.A.; Baker, E.; Sutherland, G.R.; McCaughan, G.W.
Immunogenetics 40, 331-338, 1994
A;Title: Genomic organization, exact localization, and tissue expression of the human CD
A;Reference number: S59857; MUID:95012454; PMID:7927537
A;Accession: S59857
A;Molecule type: DNA
A;Residues: 1-436,'S',438-766 <ABB>
A;Cross-references: UNIPARC:UPI000004F7BF; EMBL:U13734
C;Genetics:
A;Gene: GDB:DDP4
A;Cross-references: GDB:125239; OMIM:102720
A;Map position: 224.3-2q24.3
A;Introns: 2/3; 32/1; 35/1; 65/1; 95/3; 122/3; 140/2; 164/3; 205/1; 258/3; 296/2; 341/3; 356/3
C;Superfamily: dipeptidyl-peptidase IV
C;Keywords: dipeptidylpeptidase hydrolase; glycoprotein; homodimer; proteinase; transmembr
F;1-6/Domain: intracellular #status predicted <INT>
F;7-28/Domain: transmembrane #status predicted <EXT>
F;29-766/Domain: extracellular #status predicted <EXT>
F;85,92,150,219,229,281,321,520,685/Binding site: carboxylate (Asn) (covalent) #status
F;630,708,740/Active site: Ser, Asp, His #status predicted
Query Match 24.3%; Score 981.5; DB 1; Length 766;
Best Local Similarity 32.5%; Pred. No. 8.7e-56;
Matches 249; Conservative 131; Mismatches 304; Indels 83; Gaps 27;
QY 11 TGGSKRLTNETVTKQAITPTSRVQWLSGADGSGVVAARPGSLTIENIVNRSRTLI 70
DB 36 TADSRKTYLTLD-YLNKTYRLKLYSLRWIS---DHELYLKQENNLVFNAYGNSSVFLE 91
QY 71 PADKIPGKAENYVHPDLSSVLWASNHTKQYRHSFPADYVQDVESLKSVPIMPDOE- 129
DB 92 NSTFDEFGHINDYISPDQOFILLENNYKQWRHSYASY---DIYDLNKRQLITEERI 148
QY 130 -GDIQVQAQNSPGVNTIAFVRENDLYV--WNGTWTTRITDDGGPD-MFHGVPDMDIYBEIL 185
DB 149 PNNTQVWTFSPVGHKLAAYVNNDIYKIEPNLPSYRITWTGKEDIYNGITDMDIYBEVF 208
QY 186 GDRVALWSPDGEVLAYLSFNETGVPYTYVQYVNDQGEIAPAYPWELKIRYKVSQNTPT 245
DB 209 SAYSLAWSPNGTFLAYAQFNDTEVPLIEYSFYSDE---SLQVPKTYRVVPYKAGAVNPT 265
QY 246 VTLSLNI--ASKEVKQAPIDAPESTDLIIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297
DB 266 VKFVVNTDLSVSTNATSIQTAPASMLGDHVLCDVTWATQER--ISLQWLRRIQNS 323

RESULT 4

S23752
dipeptidyl-peptidase IV (EC 3.4.14.5) alpha chain - mouse
N;Alternate names: CD26 alpha subunit; THAM alpha subunit
C;Species: Mus musculus (house mouse)
C;Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
C;Accession: S23752; A46455; A56030
R;Marguet, D.; Bernard, A.M.; Vivier, I.; Darmoul, D.; Naquet, P.; Pierres, M.
J. Biol. Chem. 267, 2200-2208, 1992
A;Title: cDNA cloning for mouse thymocyte-activating molecule. A multifunctional ecto-d
A;Reference number: S23752; MUID:92129288; PMID:1370813
A;Accession: S23752
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-760 <MAR>
A;Cross-references: UNIPARC:UPI0000172A2D; EMBL:X58384
R;Vivier, I.; Marguet, D.; Naquet, P.; Bonicel, J.; Black, D.; Li, C.X.; Bernard, A.M.;
J. Immunol. 147, 447-454, 1991
A;Title: Evidence that thymocyte-activating molecule is mouse CD26 (dipeptidyl peptidase
A;Reference number: A46465; MUID:91302787; PMID:1712807
A;Accession: A46465
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-20 <VIV>
A;Cross-references: UNIPARC:UPI0000172A2E
A;Experimental source: M14.T thymoma cells, Swiss nu/nu
A;Note: sequence extracted from NCBI backbone (NCBI:42236)
R;Bernard, A.M.; Mattei, M.G.; Pierres, M.; Marguet, D.
Biochemistry 33, 15204-15214, 1994
A;Title: Structure of the mouse dipeptidyl peptidase IV (CD26) gene.
A;Reference number: A56030; MUID:95092780; PMID:7999781
A;Accession: A56030
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 746-760 <BER>
A;Cross-references: UNIPARC:UPI0000172A2F; GB:U12620
C;Genetics:
A;Gene: CD26
C;Superfamily: dipeptidyl-peptidase IV

C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein
F:213,223,315,514,679/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:624,702,734/Active site: Ser, Asp, His #status predicted

```
Query Match      24.3%; Score 981; DB 1; Length 760;
Best Local Similarity 33.0%; Pred. No. 9.3e-56;
Matches 251; Conservative 127; Mismatches 281; Indels 102; Gaps 29;

QY 25 VKQAITPTSRVSQVLSGAEDGSYVVAEDGSLTI-----ENIVTNEGRSLIIPADKIPTGK 79
DB 25 VKQAITPTSRVSQVLSGAEDGSYVVAEDGSLTI-----ENIVTNEGRSLIIPADKIPTGK 79
QY 47 LKSTFRVKSYSLSWVVS---DREYLYKQENNILLNNAEKGNSISFLENSTF-----93
DB 47 LKSTFRVKSYSLSWVVS---DREYLYKQENNILLNNAEKGNSISFLENSTF-----93
QY 80 EAFNYW-IHPDLSSVLMASNHTKQYRHSFFADYVQDVESLSKSVLMPDQE--GDIQVAQ 136
DB 80 EAFNYW-IHPDLSSVLMASNHTKQYRHSFFADYVQDVESLSKSVLMPDQE--GDIQVAQ 136
QY 94 ESFGVHSVSPDLFLVLLVYVNVKQWHSYASNYIDVNRKQ---LITEEKIPNNTQWIT 150
DB 94 ESFGVHSVSPDLFLVLLVYVNVKQWHSYASNYIDVNRKQ---LITEEKIPNNTQWIT 150
QY 137 WSPGVNITAFVRENDLYVDNGTVT--RIITDDGPD-MFHGVPDVIYEEELIDRYALWF 193
DB 137 WSPGVNITAFVRENDLYVDNGTVT--RIITDDGPD-MFHGVPDVIYEEELIDRYALWF 193
QY 151 WSPGHLKLAIVKNDIYVYKVPFLPSPHRSRITGTGEENVLYINGITDVMVVEEVFGAYSALWM 210
DB 151 WSPGHLKLAIVKNDIYVYKVPFLPSPHRSRITGTGEENVLYINGITDVMVVEEVFGAYSALWM 210
QY 194 SPDGEYLAYLGFNETGVPTTYVQVYMDNQEIAPAPYPMELKIRYPKVQSNPTVTLSLLNI 253
DB 194 SPDGEYLAYLGFNETGVPTTYVQVYMDNQEIAPAPYPMELKIRYPKVQSNPTVTLSLLNI 253
QY 211 SPNWTFLAYAQNDGTGVLIEYSFVSDS---SLQYPKTWIPIYPKAGAVNTEVKEFVNI 267
DB 211 SPNWTFLAYAQNDGTGVLIEYSFVSDS---SLQYPKTWIPIYPKAGAVNTEVKEFVNI 267
QY 254 --ASKEVKQAPID-----AFESTDLIIGEVAWLIDHTTVAAKAFNRVQDQKVAVDVA 306
DB 254 --ASKEVKQAPID-----AFESTDLIIGEVAWLIDHTTVAAKAFNRVQDQKVAVDVA 306
QY 268 DSLSSSSAAPIQIPAPASVARGDHYLDCVVMATEER--ISLQMLRRITQNSVMAICDYD 325
DB 268 DSLSSSSAAPIQIPAPASVARGDHYLDCVVMATEER--ISLQMLRRITQNSVMAICDYD 325
QY 307 SNKAT--VISRSD--CTDGLNDLLSMKYIGPIKP-----SDKDAYYIDIDSHGWA 354
DB 307 SNKAT--VISRSD--CTDGLNDLLSMKYIGPIKP-----SDKDAYYIDIDSHGWA 354
QY 326 KINLTWNCPSQOQHVEMSTTG-----VGRFRPAEPHFTSDGGSFYKILSDXGYK 376
DB 326 KINLTWNCPSQOQHVEMSTTG-----VGRFRPAEPHFTSDGGSFYKILSDXGYK 376
QY 355 HLYLFPVSGGPIPTKGDWEVTSILSDQERQLVYLYLSTQHHSTE--RHLYSYSYSTFA 412
DB 355 HLYLFPVSGGPIPTKGDWEVTSILSDQERQLVYLYLSTQHHSTE--RHLYSYSYSTFA 412
QY 377 HICHPFKDKKCTFTKGAWEVISEALTSD--YLYISNQYKEMPGGRNLYKIQ-----429
DB 377 HICHPFKDKKCTFTKGAWEVISEALTSD--YLYISNQYKEMPGGRNLYKIQ-----429
QY 413 VTPLVDDT-----VAAVWSAIFSANSYGYIITYGDPVPYQELVYTNSTKPLR 460
DB 413 VTPLVDDT-----VAAVWSAIFSANSYGYIITYGDPVPYQELVYTNSTKPLR 460
QY 430 ---LTDHTNVKCLSDLNPERCQYVAVSFSKEAKYQLGCGVGLPLTLHRSIDHKELR 486
DB 430 ---LTDHTNVKCLSDLNPERCQYVAVSFSKEAKYQLGCGVGLPLTLHRSIDHKELR 486
QY 461 TITDNAKVLBQIKDYALPN--ITYEFLPLPSGETLNVNQR-LPPGFSDDKYPILFTPYG 517
DB 461 TITDNAKVLBQIKDYALPN--ITYEFLPLPSGETLNVNQR-LPPGFSDDKYPILFTPYG 517
QY 487 VLEDNSALDMLQDQVMSKCLDFTVL-----NETRFWYQMLTPPHFDKSKYPLLLDYYA 542
DB 487 VLEDNSALDMLQDQVMSKCLDFTVL-----NETRFWYQMLTPPHFDKSKYPLLLDYYA 542
QY 518 GFGAEVTKRMQALNFKAVVASDSLELYVTWVNRGTGFGKGRKFRSAVTROLGLLEAED 577
DB 518 GFGAEVTKRMQALNFKAVVASDSLELYVTWVNRGTGFGKGRKFRSAVTROLGLLEAED 577
QY 543 GPCSKRAADASFRLNWAYVLAETENI--IVASFQGRGSGYQGDXTMHAINRLGLTLEVED 599
DB 543 GPCSKRAADASFRLNWAYVLAETENI--IVASFQGRGSGYQGDXTMHAINRLGLTLEVED 599
QY 578 QIYAAQQAANIPWIDADHIGWGSFGGYLTSKVLKDSGAPTLGVITAPVSDMRFYDSM 637
DB 578 QIYAAQQAANIPWIDADHIGWGSFGGYLTSKVLKDSGAPTLGVITAPVSDMRFYDSM 637
QY 600 QLEEARQFVGMGFVDSKRAVIMGWSYGGVYVTSVMVLGSGGVFPKCGIAVAPVSRMEYDSV 659
DB 600 QLEEARQFVGMGFVDSKRAVIMGWSYGGVYVTSVMVLGSGGVFPKCGIAVAPVSRMEYDSV 659
QY 638 YTERYM--KLTSTNEGVEYTSVAV-RKTDGFKVVEGGFLIOHGTGDDNVHFQNSAALVDLL 694
DB 638 YTERYM--KLTSTNEGVEYTSVAV-RKTDGFKVVEGGFLIOHGTGDDNVHFQNSAALVDLL 694
QY 660 YTERYMGFLIPEDNLNDLHVRNSTMVSRAEHFQKQVE--YLLIHGTADNDNVHFQOQSAISKVL 717
DB 660 YTERYMGFLIPEDNLNDLHVRNSTMVSRAEHFQKQVE--YLLIHGTADNDNVHFQOQSAISKVL 717
QY 695 MGDGVSPKELHSGQWFTSDPHGYSYHGGVFLYKQLARKLYQ 735
DB 695 MGDGVSPKELHSGQWFTSDPHGYSYHGGVFLYKQLARKLYQ 735
QY 718 VDAGVD---FQAWWYIDEDHGASSTAHQHIYSHMSHFLOQ 755
DB 718 VDAGVD---FQAWWYIDEDHGASSTAHQHIYSHMSHFLOQ 755
```

RESULT 5
A:Accession: A39914
A:Molecule type: mRNA
A:Residues: 1-792 <HON>
A:Cross-references: UNIPROT:P14740; UNIPARC:UPI000017098A; GB:J02997; NID:g204463; PIDN:
J:Ogata, S.; Misumi, Y.; Ikehara, Y.
R: Biol. Chem. 264, 3596-3601, 1989
A:Title: Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA an
A:Reference number: A33315; MUID:89123496; PMID:2563382
A:Accession: A33315
A:Molecule type: mRNA
A:Residues: 1-37, 'A', '39-182, 'I', '184-331, 'T', '333-351, 'C', '353-393, 'V', '395-561, 'L', '563-623,
A:Cross-references: UNIPARC:UPI0000129841; GB:J04591; NID:g203973; PIDN:AAA41096.1; PID:
A:Note: the authors translated the codon GCG for residue 38 as Arg, ACC for residue 332
A:Accession: B33315
A:Molecule type: protein
A:Residues: 1-20, '35-54, '427-443, '505-509, '511-520, '530-538, '593-600, '602-608, '618-627 <HO2>
A:Cross-references: UNIPARC:UPI0000172A2E; UNIPARC:UPI0000172A30; UNIPARC:UPI0000172A31;
A36: UNIPARC:UPI0000172A37
R:McCaughan, G.W.; Wickson, J.E.; Creswick, P.F.; Gorrell, M.D.
Hepatology 11, 534-544, 1990
A:Title: Identification of the bile canalicular cell surface molecule GP110 as the ectop
quence.
A:Reference number: A60730; MUID:90228896; PMID:1970322
A:Accession: A60730
A:Molecule type: protein
A:Residues: 28-47, 'XX', '50-53, '55-58 <MC>
A:Cross-references: UNIPARC:UPI0000172A38
R:Ogata, S.; Misumi, Y.; Tsuji, E.; Takami, N.; Oda, K.; Ikehara, Y.
Biochemistry 31, 2582-2587, 1992
A:Title: Identification of the active site residues in dipeptidyl peptidase IV by affini
A:Reference number: A42203; MUID:92190188; PMID:1347701
A:Accession: A42203
A:Molecule type: protein
A:Residues: 'R', '625-630, 'X', '632-648 <OG2>
A:Cross-references: UNIPARC:UPI0000172A39
R:Iwaki-Egawa, S.; Watanabe, Y.; Fujimoto, Y.
Biol. Chem. Hoppe-Seyler 374, 973-975, 1993
A:Title: N-terminal amino acid sequence of the 60-kDa protein of rat kidney dipeptidyl
A:Reference number: S38949; MUID:94128239; PMID:7905271
A:Accession: S38949
A:Status: preliminary
A:Molecule type: protein
A:Residues: 281-302 <IWA>
A:Cross-references: UNIPARC:UPI0000172A3A
R:Hong, W.; Doyle, D.
J. Biol. Chem. 263, 16892-16898, 1988
A:Title: Membrane orientation of rat gp110 as studied by in vitro translation.
A:Reference number: A31781; MUID:89034185; PMID:3182821
A:Accession: A31781
A:Molecule type: mRNA
A:Residues: 1-40 <HO3>
A:Cross-references: UNIPARC:UPI0000172A3B
C:Comment: This protein is localized to the bile canalculus, which is the apical domain
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidyl-peptide hydrolase; glycoprotein; homodimer; liver; serine proteina
F:1-792/Product: dipeptidyl-peptidase, membrane-bound form #status experimental <MATM>
F:1-28/Domains: signal sequence #link MATS #status experimental <SIG>
F:1-6/Domains: intracellular #status predicted <INT>
F:7-28/Domains: transmembrane #status predicted <TMN>
F:29-792/Domains: extracellular #status predicted <EXT>
F:29-34/Domains: propeptide #link MATS #status experimental <PRO>
F:35-792/Product: dipeptidyl-peptidase, soluble form #status experimental <MATS>
F:83,90,148,217,227,319,521,686/Binding site: carbohydrate (Asn) (covalent) #status pre
F:631/Active site: Ser #status experimental
F:709,741/Active site: Asp, His #status predicted

Query Match 23.5%; Score 949.5; DB 1; Length 792;
Best Local Similarity 32.2%; Pred. No. 1.1e-53;
Matches 245; Conservative 129; Mismatches 292; Indels 95; Gaps 27;
QY 25 VKQAITPTSRVSQVLSGAEDGSYVVAEDGSLTIENIVTNEGRSLIIPADKIPTGKEAFNY 84
DB 47 LKNTFRVKSYSLSRWVS---DSEYLYKQENNILLNNAEKGNSISFLENSTFTEIFGDSISDY 103

submitted to the EMBL Data Library, October 1995
A;Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV
A;Reference number: S60938
A;Accession: S60946
A;Molecule type: DNA
A;Residues: 1-931 <GAL>
A;Cross-references: UNIPARC:UPI0000136060; EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; F
R;Boyer, J.; Fairhead, C.; Gallion, L.; Gallison, F.; Michaux, G.; Thierry, A.; Dujon, B.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67104
A;Accession: S67112
A;Molecule type: DNA
A;Residues: 1-931 <BOY>
A;Cross-references: UNIPARC:UPI0000136060; EMBL:Z75127; NID:g1420507; PIDN:CAA99437.1; F
R;Gallison, F.; Dujon, B.
Yeast 12, 877-885, 1996
A;Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV
A;Reference number: S71713; MUID:96437977; PMID:8840505
A;Accession: S71721
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-931 <GAW>
A;Cross-references: UNIPARC:UPI0000136060; EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; F
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C;Genetics:
A;Gene: SGD:STE13; YC11
A;Cross-references: SGD:S0005745; MIPS:YOR219C
A;Map position: 15R
C;Function:
A;Description: involved in processing of alpha-factor prepropheromone
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein
F:113-150/Domain: transmembrane #status predicted <TM>
F:377/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:785,863,896/Active site: Ser, Asp, His #status predicted

Query Match 22.4%; Score 902.5; DB 2; Length 931;
Best Local Similarity 32.2%; Pred. No. 1.6e-50;
Matches 230; Conservative 134; Mismatches 265; Indels 85; Gaps 25;

QY 86 IHPDLSVLWASNHTKQYRHSFFADYYVQDVESLSKSVPLMPDQEGP-----LQYAQ 136
DB 246 INYKDKLFGTNLESEFRHSKGFYWKDLANTGNTEPILPPEKSDNDYELGSLKLSYAH 305
QY 137 WSPVGNVTIAFVRENDLYVM--DNGTVTRITDGGDPMHFGVDPWIYEEETILGDRYALWFS 194
DB 306 FSPAYNYIFVTENNLFLOQVNSGVAKKVTEDGSKDIFNAKPDWYIEEVEVLASDQALWVA 365
QY 195 PDGEVLAFLSNFNETGVPTTYVQYMDNQBIAPAYPMELKIRYKVSQTNPTVTLNLNIA 254
DB 366 PDDSKAVFARFNDTSVDDIRLNRYTNMNE--AYLSDTKIKYKPGFQNPQPDFLVNLQ 422
QY 255 SKEVQAPIDAFESTDLIIIGEVAWLT-DTHTTVAAKAFNRVQDQKQVAVDTASNKATVI 313
DB 423 NGIYS--INTGGQKDSILYNGKWSIPDFRFEITDRNSKILD---VKYVDIPSSQMLTV 477
QY 314 SDRDGT--DGWLD---NLASMKYIGIKPSDK--DAYYIDI-SDHSGWAHLVLP--VSGG 364
DB 478 RTNNSNLFNGWIEKTKDILSI---PPKPELKMDDYGYDIHADSGFSLFYPTVFAK 533
QY 365 EPIPLTKGWDEVT--SILSIDORQLVYYLSTOHHSTERHLXSVSYSTFAVTPLVDDTVA 422
DB 534 EPIQLTKGNWETNGVGVYEYETDIFFTANEIGVMSQHLXIS-----LTDSTTQ 585
QY 423 -----AYWSAFSANSYGVIITYGGDPVDPYQ-----ELYTNSTK 457
DB 586 NTFQSLQNPDKYDFYDFELSSSARYAISKCLGPDTPFKVAGPLTRVLNVAETHDS--- 642
QY 458 PLRTITDNKVLKQIKDYALPNITYFELPLPSETLNVMQRLPPGSPDKKYPIILFTPYG 517
DB 643 -ILQLTKDEKFEKIKNYDLPITSYKTMVLDGVEINYIEIKPANLNPKKYPIILVNIY 701
QY 518 GPGAEVTKRWQALNFKAYVASDSELEYVTWTVNDRGTGKGRKFRSAVTRQLGLEAED 577

RESULT 6
A49737
dipeptidyl aminopeptidase (EC 3.4.14.-) - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein O5045; protein YOR219c; protein YOR50-9
C;Species: Saccharomyces cerevisiae
C;Date: 03-Mar-1994 #sequence revision 03-Mar-1994 #text_change 09-Jul-2004
C;Accession: A49737; S45451; S60946; S67112; S71721
R;Santa Anna-A, S.; Herskowitz, I.
A;Title: Isolation and DNA sequence of the STE13 gene encoding dipeptidyl aminopeptidase
submitted to the Protein Sequence Database, July 1993
A;Reference number: A49737
A;Accession: A49737
A;Molecule type: DNA
A;Residues: 1-931 <SAN>
A;Cross-references: UNIPROT:P33894; UNIPARC:UPI0000136060; GB:L21944; NID:g347196; PIDN:
R;Anna-Arriola, S.S.; Herskowitz, I.
Yeast 10, 801-810, 1994
A;Reference number: S45451; MUID:95066382; PMID:7975897
A;Accession: S45451
A;Molecule type: DNA
A;Residues: 1-931 <ANN>
A;Cross-references: UNIPARC:UPI0000136060; EMBL:L21944; NID:g347196; PIDN:AAA35119.1; PI
R;Gallison, F.; Dujon, B.

85 WIHPDLSVLWASNHTKQYRHSFFADYYVQDVESLSKSVPLMPDQEGP-----LQYAQWSPVGN 142
DB 104 SVSPDRFLVLEYNVYKQWHSYASYSD---LNKRQLITEKIPNNQTMTWQSEGH 160
QY 143 TIAFVRENDLYWMDNGTVT--RLTDDGGDP-MFHGVDPMYIEEETILGDRYALWFSDPGEY 199
DB 161 KLAYVWKNDIYKIEPHLPSHRTTSGKENVIFNGINDWYIEEIEFGAYSALWWSPTGTF 220
QY 200 LAYLSNFGNETGVPTTYVQYMDNQBIAPAYPMELKIRYKVSQTNPTVTLNLNIAKSEVK 259
DB 221 LAYAQFNDTGVPLEIYFSYDSE---SLQYPKTVMIPYKAGAVNPVKFIVNTDLSST 277
QY 260 QAPID-----AFESTDLIIIGEVAWLT-DTHTTVAAKAFNRVQDQKQVAVDTASNKATVI 312
DB 278 TITPMTQITAPASVTGGDHYLCDAVMSDR--ISLQRLRRIQNSVMAICD--YDKTNL 333
QY 313 ISDRDGT-----GWLNDLSMKYIGIKP-----SKDAYYIDI-SDHSGWAHLVYL 358
DB 334 VMCPTTQEHETSATGW-----VGRFRPAEPHFTSDGSSFYKIVSDKDGKXHCICQ 384
QY 359 FVPSGGEPIP---LTKGWDEVTLSILSIDORQLVYLSLQHHSTE--RHLXSVSYSTFA 412
DB 385 FQ-KDRKPEODCTFITKGAWEVISIEALTSD--YLYISNEYKEMPGGRNLYKIQ----- 436
QY 413 VTPLVDDT-----VAAWMSAFSANSYGVIITYGGDPVYPQELYYTNTSKPLR 460
DB 437 ---LTHTNKKCLSCDLNPERCQYYSVLSKEAKYQLGCRGPGPLPLYTLHRSTDQKELR 493
QY 461 TITDNKVLKQIKDYALPN--ITYFELPLPSGTTLNMQR-LPPGSPDKKYPIILFTPYG 517
DB 494 VLEDNSALDKMLQDVQMPKSLDFVL---NETREWYQMLPLPHFDKSKYKPLLLIDVYA 549
QY 518 GPGAEVTKRWQALNFKAYVASDSELEYVTWTVNDRGTGKGRKFRSAVTRQLGLEAED 577
DB 550 GPCSQKADAAFR-FNWATYLASTENI--IVASPDGRSGYQGDKIMHAINKRLGLTLEVED 606
QY 578 QIVAAQQAANIPWIDADHIGIWCSPGGLTSLVKLEKSGAFTLVITAPVSDWRFYDSM 637
DB 607 QIEAARQFLKMGFVDSKQVAINGWSYGGVYTSNVLGSGGVFKCGIATAPVSEWEYDSV 666
QY 638 YTERYM--KTLSTNEBGEYTSYAV-RKTDGFKNVVEGFLIOHGTGDDNVHVFQNSAALVLL 694
DB 667 YTERYMGLPTEPNDLHRYNSTVMSRAENPKQVE--YLLIHGTADDNVHVFQNSAQSKAL 724
QY 695 MGDGVSPKLSHWPFTSDHGHSYHGGVFLYQLARKLYQ 735
DB 725 VDAGVD---FQAWMYTDEDHGIASSTAHOHIYSHMGSFLLQO 762

submitted to the EMBL Data Library, October 1995
A;Description: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV
A;Reference number: S60938
A;Accession: S60946
A;Molecule type: DNA
A;Residues: 1-931 <GAL>
A;Cross-references: UNIPARC:UPI0000136060; EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; F
R;Boyer, J.; Fairhead, C.; Gallion, L.; Gallison, F.; Michaux, G.; Thierry, A.; Dujon, B.
submitted to the Protein Sequence Database, July 1996
A;Reference number: S67104
A;Accession: S67112
A;Molecule type: DNA
A;Residues: 1-931 <BOY>
A;Cross-references: UNIPARC:UPI0000136060; EMBL:Z75127; NID:g1420507; PIDN:CAA99437.1; F
R;Gallison, F.; Dujon, B.
Yeast 12, 877-885, 1996
A;Title: Sequence and analysis of a 33 kb fragment from the right arm of chromosome XV
A;Reference number: S71713; MUID:96437977; PMID:8840505
A;Accession: S71721
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-931 <GAW>
A;Cross-references: UNIPARC:UPI0000136060; EMBL:X92441; NID:g1050762; PIDN:CAA63182.1; F
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1995
C;Genetics:
A;Gene: SGD:STE13; YC11
A;Cross-references: SGD:S0005745; MIPS:YOR219C
A;Map position: 15R
C;Function:
A;Description: involved in processing of alpha-factor prepropheromone
C;Keywords: dipeptidylpeptide hydrolase; glycoprotein; transmembrane protein
F:113-150/Domain: transmembrane #status predicted <TM>
F:377/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:785,863,896/Active site: Ser, Asp, His #status predicted

Query Match 22.4%; Score 902.5; DB 2; Length 931;
Best Local Similarity 32.2%; Pred. No. 1.6e-50;
Matches 230; Conservative 134; Mismatches 265; Indels 85; Gaps 25;

QY 86 IHPDLSVLWASNHTKQYRHSFFADYYVQDVESLSKSVPLMPDQEGP-----LQYAQ 136
DB 246 INYKDKLFGTNLESEFRHSKGFYWKDLANTGNTEPILPPEKSDNDYELGSLKLSYAH 305
QY 137 WSPVGNVTIAFVRENDLYVM--DNGTVTRITDGGDPMHFGVDPWIYEEETILGDRYALWFS 194
DB 306 FSPAYNYIFVTENNLFLOQVNSGVAKKVTEDGSKDIFNAKPDWYIEEVEVLASDQALWVA 365
QY 195 PDGEVLAFLSNFNETGVPTTYVQYMDNQBIAPAYPMELKIRYKVSQTNPTVTLNLNIA 254
DB 366 PDDSKAVFARFNDTSVDDIRLNRYTNMNE--AYLSDTKIKYKPGFQNPQPDFLVNLQ 422
QY 255 SKEVQAPIDAFESTDLIIIGEVAWLT-DTHTTVAAKAFNRVQDQKQVAVDTASNKATVI 313
DB 423 NGIYS--INTGGQKDSILYNGKWSIPDFRFEITDRNSKILD---VKYVDIPSSQMLTV 477
QY 314 SDRDGT--DGWLD---NLASMKYIGIKPSDK--DAYYIDI-SDHSGWAHLVLP--VSGG 364
DB 478 RTNNSNLFNGWIEKTKDILSI---PPKPELKMDDYGYDIHADSGFSLFYPTVFAK 533
QY 365 EPIPLTKGWDEVT--SILSIDORQLVYYLSTOHHSTERHLXSVSYSTFAVTPLVDDTVA 422
DB 534 EPIQLTKGNWETNGVGVYEYETDIFFTANEIGVMSQHLXIS-----LTDSTTQ 585
QY 423 -----AYWSAFSANSYGVIITYGGDPVDPYQ-----ELYTNSTK 457
DB 586 NTFQSLQNPDKYDFYDFELSSSARYAISKCLGPDTPFKVAGPLTRVLNVAETHDS--- 642
QY 458 PLRTITDNKVLKQIKDYALPNITYFELPLPSETLNVMQRLPPGSPDKKYPIILFTPYG 517
DB 643 -ILQLTKDEKFEKIKNYDLPITSYKTMVLDGVEINYIEIKPANLNPKKYPIILVNIY 701
QY 518 GPGAEVTKRWQALNFKAYVASDSELEYVTWTVNDRGTGKGRKFRSAVTRQLGLEAED 577

418 YPGRNIVRISIGSVPPSPKCVTCHLRKRCQYYTASFSDYAKYVYALVCYGGPIPISTLH 477
 452 TTNSTKPLRTITDIAKVLQKDYALPNITFELPLPSGETLNVQRLPPGSPDKKYP1 511
 478 DGRDQDEIKILEENKELNALKNITQLPKERTKKLEVD-ITLWYKMLPPQFDRSKYPL 536
 512 LFTPYGGCAQEVTKRQALNFKAVASDSLEYVVTWVNRGCTGFGKGRKFRSAVTRQIG 571
 537 LIQVYGGPCSSVRVSF-AVNWISYLASKGM--VIALVDGRGTAFOGDKLLIYAYRKL 593
 572 LLEAEDQIYAAQAAANIPIWDADHIGIWNVSGFGYLTISKLEKSGAFTLGLVITAP-VSD 630
 594 VYVEDQITAVRKFTIEMGFIDEKRIALNGWS-----YEIRFITGPCIWN 637
 631 WRF-----YDS-----MYTERY-MKTLSTNEEGYETSAV-RKTDGFKNVGGFLI 673
 638 WSFQWYSGSSLQILGLRHLRHHGSPNKDDNLEHYKGNSTVWARAEYFRNVD--YLL 695
 674 QHGTGDNVHFNSAALVDLLMGDGVSPKHLHQWFTSDHGIS 717
 696 IHGTADDNVHFNSAQIAKALVNAQVD--FQAMWYSDQNHLG 736

 RESULT 8
 S6261
 X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Flavobacterium meningosepticum
 C:Species: Flavobacterium meningosepticum
 C:date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
 C:Accession: S6261
 R:Kabashima, T.; Yoshida, T.; Ito, K.; Yoshimoto, T.
 Arch. Biochem. Biophys. 320, 123-128, 1995
 A:Title: Cloning, sequencing, and expression of the dipeptidyl peptidase IV gene
 A:Reference number: S6261, MUID:95314307; PMID:7793970
 A:Accession: S6261
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-711 <RAB>
 A:Cross-references: UNIPROT:Q47900; UNIPARC:UPI00000B2115; EMBL:D42121; NID:G57
 C:Superfamily: dipeptidyl-peptidase IV
 C:Keywords: dipeptidylpeptide hydrolase

 Query Match 20.1%; Score 811.5; DB 2; Length 711;
 Best Local Similarity 30.0%; Pred. No. 8.7e-45;
 Matches 207; Conservative 135; Mismatches 278; Indels 71; Gaps 20;

 QY 46 SYVYAAEDGSLTIENIVTNSRTLIPADKIPTGKEAFN-----PTGIKYSYKTSQKEKNIVDGSFGQYFSN 85
 DB 40 SGIASLNDG-----ENYATIE-----
 QY 89 DLSSVLWASNHTKQYRHSFFADYVYQVESLKSVPIMPDOEGDIQYAOQWSVGNITAFVR 148
 DB 86 DESKILLQKSSQSYRHSFLGKFEVKOLKSRVTVSL--NNANWIQEPKFSPDGSKVAFIA 143
 QY 149 ENDLYVVD--NGVTRITDDGGP--DMFHGVPDWIYEEBELLGDRYALMFSPOGEYLAYLSF 205
 DB 144 DNNLFYQDLNTGKTIQTITDGKKNEINGLDGWYEESEFGHADYYQW-NKAGDALVFRVF 202
 QY 206 NETGVPTVYQYMDNQBIAIPAYFWELKIRYPKVSQNTPTVTLNLNTASKEVQAQIDA 265
 DB 203 DERKVEINIIYYQN-----LYPKLMTYKYPKAGEENSAVTAVLYQLSSG--KSAQLNF 255
 QY 266 FESTDLIIGEVAWLTDDHTTVAAKAFNRVQOQKVAVDVTASNKATVLSID--RQDGDGWL 323
 DB 256 GSSEKYYIQL-FQTNANDEIVVATANKHONKVDLLKVT---KTAAYSKLFTETDNAMI 311
 QY 324 --DNLLSMKYIGPIKPSDKDAYIDI SDHSGWAHLVLPFVSGGEBIPITKGDWEVTSILS 381
 DB 312 EFTD-LTWEFL-----DDNSFLWASRDEGRHLLYWDYDAGLKKQVSKGDWEIINYG 363
 QY 382 IDQERQLVYLTSTQHSTHERHLYSVSYFTAVTPLVDDTVAAVWYASFSANSGYLLTVG 441
 DB 364 YNPKTEVYIQTTEKGSINKVVSKLNIINT-GKTQLLSNAGNN-SAAFSKTFNYFINTSS 421

```
QY 442 GPDVPYQELYYTNTSKPLRTITDIAKVLQIKDYALPNITYFELPLPSGETLNNMQLRPP 501
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 422 TAKVPTKYILKDANGKVKELQNDDLLNKLKSNDFIAKEFITIPNAAGDQNNAMWIKPK 481

QY 502 GPSPKKYPILFTPYGPGAEVETKRWQALNPKAYVASDSLEYVYTWVDRGTGFKGRK 561
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 482 NFDPAKKYVFMFQYSGPGSQVANSWDGCGN-GIWPDLMAQKGYLVVCDGRGTGFRGTK 540

QY 562 FRSAVTRQGLLEAEADQIYAAQQAANI PWIDADHIGIWMGSGGYLTSTKVLKDSGAFTL 621
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 541 YKKVTYKNLKGKYEIDQITAAKWLGNQSYVDSKRGIFGWSYGGVNASLMTKGADEVFKK 600

QY 622 GVITAPVSDWRPFYDSMYTERYMKTLSTNEEGYETSAVRKTDGFKVVEGFLIQHGTGDDN 681
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 601 GIAPVPTVNRVYDSIYTERFLOTPOENKQGVDLNS--PTTVAKLLKGRKFLIHGTADDN 658

QY 682 VHFQNSAALVDLLMGDG-----VSPKHLHS 706
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 659 VHFQNSMEPSEALIQNKQDFMAYDPKNHS 689

RESULT 9
168600
diptidyl aminopeptidase like protein - human
C:Species: Homo sapiens (man)
C>Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 21-Jul-2000
C:Accession: 168600
R:Yokotani, N.; Doi, K.; Wenthhold, R.J.; Wada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A:Title: Non-conservation of a catalytic residue in a diptidyl aminopeptidase IV-rela
A:Reference number: I54331; MUID:93372805; PMID:8103397
A:Accession: 168600
A:Status: preliminary; translated from GB/EMBL/DBSJ
A:Molecule type: mRNA
A:Residues: 1-803 <RES>
A:Cross-references: UNIPARC:UPI000016A7F9; GB:M96860; NID:G306707; PIDN:AAA35761.1; PID:
C:Superfamily: diptidyl-peptidase IV

Query Match 19.3%; Score 779.5; DB 2; Length 803;
Best Local Similarity 29.5%; Pred. No. 1.3e-42;
Matches 226; Conservative 126; Mismatches 318; Indels 95; Gaps 31;

QY 37 QWLSGAEDGSYVYAAEDGSLTIENIVTNESRTLI PADKIPTGKEAFNYIHPDLSSLVLA 96
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 87 KWS---DTEFIYREQGTVRLNVTNTSTVLEGGKIES-LRAIRYEISDPREYALFS 142

QY 97 SNHTKOYRHSFFADYTVQDVESLSKSVPLMPDQEGD-----IQYAOHSPVGNITAFVR 148
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 143 YNVEPIYQHS-YTGYVY-----LSKIPHGDPQSLDPPEVSNAKLQYAGWPKGQQLIFIP 196

QY 149 ENDLYVWDN--GTVTRITDDGGPD-MFHGVPDWIYEEIILGDRYALWFGPDGEYLAYLSF 205
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 197 ENNIYCAHVGAQKAIKRVSTGKEGVYNGLSDWLYEEIILKTHIAHWSPDGTGLAYAAI 256

QY 206 NETGV-----PTVTQVYMDNQBIAPAYPWEIKIRYPKVSQNTPTVTLSL--LNIAKSEV 258
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 257 NDSRVPIMELPTIT-----GSIYPTVKPYHPKAGSENPISLSLHVIGLNGTHDL 306

QY 259 KOAPIDAFESTDLIIGEVAWLTDTHTTVAAKAFNRVQDQKQVAVDTASNKATVISDRDG 318
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 307 EMMPDDPRMREYIITWVKWATS--TKVAVTMLNRAQNVSIILTCDAITGVCCTK-KHEDE 363

QY 319 TDGWLNLDSMKYIGIPKPSD-KDYYIDISDHSQWALHYLPFVSGGEP-----IPLT 370
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 364 SEAWLHR---QNEEPVFSKGRKFFIRAIPOGGRGKGFYHITPVSSSQPNSSNDNIQSIT 419

QY 371 KGWETVTSILSDOERQVLYLSTQHSSTERHLYSV-SYSTFAVTPLVDDTV--AAWWSA 427
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 420 SGDWDTVKILAYDEKGNKIYFLSTEDLPRRQQLYSANTEGNFRNQCLSCDLVENCYFSA 479

QY 428 SFSANSQYILTYGGPDVPYQELYYTNTSKPLRTITDIAKVLQIKDYALPNITYFELPL 487
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 480 SFSHSDMDFLLKCEGCGVPMVTVHNTTDDKKKMPDLETNEHVKKAINDRQMPKVEYRDIEI 539
```

```
QY 488 PSGETLNV-MORLPPG-FSPDKKYPILFTPYGPGAEVETK-----WQALNPKAYVASDS 541
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 540 ---DDYNLPMQLLKPAFTDTTHYPLLLVVDGTGPGSQSVAEKFEVSWETVMVSSHGA--- 593

QY 542 ELEYVTWVDRGTGFKGRKFRSAVTRQGLLEAEADQIYAAQQAANI PWIDADHIGIWMG 601
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 594 ----VVVKCDGRGSGFGQTKLLHVEVRRRLGLEERQDQMEAVRTMLKEQYIDRTVAVFGK 649

QY 602 SPGGYLTSTKVL-----EKDSGAFTLGVITAPVSDWRPFYDSMYTERYMKTLSTNEEGYE-TS 656
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 650 DVGGYLSTIILPAKENOGQOTFCGSALSPYTDFKLYASAFSERVYLGHLGNRAVEMTK 709

QY 657 AVRKTGDFKNVEGGLIQHGTGDDNVHFNQNSAALV-DLLMGDG-----VSPKHLHSQWTF 710
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 710 VAHRVSALE--EQQLIIHTADEKIHFOHTAEILQTILRGKANSYLIQYPDESH--YFT 765

QY 711 DSDHGISHVGGVFLYKQLARKLYQKQRQTOVLMHQTWKDLEE 755
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 766 SS--SLKQH-----LYRSII-NFFVECFRIQDKLPTVTVTAKEDEEE 802
```

RESULT 10

```
I54331
diptidyl aminopeptidase like protein - human
C:Species: Homo sapiens (man)
C>Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C:Accession: I54331
R:Yokotani, N.; Doi, K.; Wenthhold, R.J.; Wada, K.
Hum. Mol. Genet. 2, 1037-1039, 1993
A:Title: Non-conservation of a catalytic residue in a diptidyl aminopeptidase IV-rela
A:Reference number: I54331; MUID:93372805; PMID:8103397
A:Accession: I54331
A:Status: preliminary; translated from GB/EMBL/DBSJ
A:Molecule type: mRNA
A:Residues: 1-865 <RES>
A:Cross-references: UNIPROT:P42658; UNIPARC:UPI0000047378; GB:M96859; NID:G306705; PIDN:
C:Superfamily: diptidyl-peptidase IV

Query Match 19.3%; Score 779.5; DB 2; Length 865;
Best Local Similarity 29.5%; Pred. No. 1.4e-42;
Matches 226; Conservative 126; Mismatches 318; Indels 95; Gaps 31;

QY 37 QWLSGAEDGSYVYAAEDGSLTIENIVTNESRTLI PADKIPTGKEAFNYIHPDLSSLVLA 96
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 149 KWS---DTEFIYREQGTVRLNVTNTSTVLEGGKIES-LRAIRYEISDPREYALFS 204

QY 97 SNHTKOYRHSFFADYTVQDVESLSKSVPLMPDQEGD-----IQYAOHSPVGNITAFVR 148
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 205 YNVEPIYQHS-YTGYVY-----LSKIPHGDPQSLDPPEVSNAKLQYAGWPKGQQLIFIF 258

QY 149 ENDLYVWDN--GTVTRITDDGGPD-MFHGVPDWIYEEIILGDRYALWFGPDGEYLAYLSF 205
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 259 ENNIYCAHVGAQKAIKRVSTGKEGVYNGLSDWLYEEIILKTHIAHWSPDGTGLAYAAI 318

QY 206 NETGV-----PTVTQVYMDNQBIAPAYPWEIKIRYPKVSQNTPTVTLSL--LNIAKSEV 258
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 NDSRVPIMELPTIT-----GSIYPTVKPYHPKAGSENPISLSLHVIGLNGTHDL 368

QY 259 KOAPIDAFESTDLIIGEVAWLTDTHTTVAAKAFNRVQDQKQVAVDTASNKATVISDRDG 318
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 369 EMMPDDPRMREYIITWVKWATS--TKVAVTMLNRAQNVSIILTCDAITGVCCTK-KHEDE 425

QY 319 TDGWLNLDSMKYIGIPKPSD-KDYYIDISDHSQWALHYLPFVSGGEP-----IPLT 370
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 SEAWLHR---QNEEPVFSKGRKFFIRAIPOGGRGKGFYHITPVSSSQPNSSNDNIQSIT 481

QY 371 KGWETVTSILSDOERQVLYLSTQHSSTERHLYSV-SYSTFAVTPLVDDTV--AAWWSA 427
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 482 SGDWDTVKILAYDEKGNKIYFLSTEDLPRRQQLYSANTEGNFRNQCLSCDLVENCYFSA 541

QY 428 SFSANSQYILTYGGPDVPYQELYYTNTSKPLRTITDIAKVLQIKDYALPNITYFELPL 487
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```



```
Db 355 GNQHLRYAADGKLIQAITKQDMPVIGLEVDKARVAFSASIDTPIERRLEYVSYAK- 413
QY 412 AVTPLVDDTVAAYWSAFSANSYILTYGGDPVYQELTYTNSTKPLRTITDIAKVL- 470
Db 414 POKPALTSAGGWAAKVADNGGAFAGTYSDKTPSQALYSADGRVRWTEEN-KLAEG 472
QY 471 -QIKDYA--LPNITYFELPLPSETGNTVNMQRLLPGGSPDKYPILFTPYGGGAGQEVTKR 527
Db 473 HPYWPYAANLPQEPGSLKAADGETLHYEILKPIGFDPAKKYPAIVSVYGGPHAQRVMKN 532
QY 528 WQALNFKAYVASDSLEYVTVVNRGCTGFKGRKFRSAVTROGLLEAEQIYAAQOAN 587
Db 533 WHPSERTYL-----BAGYVIFKLDNRGSGNRSKAFMRALDRKLGTVESVDQLGAKFLAS 588
QY 588 IPWIDADHIGWGSFGGYLTSSKLEKDSGAPTLGVITAPVSDMRFYDSMYTERYMKTLS 647
Db 589 QYVVDADKLGVMGWSYGGFMALMLITAENTPFKAGAAGAPPTENSLYDTATERYMGKPD 648
QY 648 TNEEGYETSAV--RKTDGFKQNVGEGFLIQHGTGDDNVHFQNSAALVDLLMGDEV-----S 700
Db 649 ENKAGYAYSDINNRIID--KLAFGSLLLHGMADDNVIPENSTRMAALQKAILPEMAMY 706
QY 701 PEKLS 706
Db 707 PGERHS 712

RESULT 13
JC5142
X-Pro dipeptidyl-peptidase (EC 3.4.14.11) - Xanthomonas maltophilia
C:Species: Xanthomonas maltophilia
C>Date: 31-Jan-1997 #sequence_revision 27-Feb-1997 #text_change 09-Jul-2004
C:Accession: JC5142
R:Kasahima, T.; Ito, K.; Yoshimoto, T.
J. Biochem. 120, 1111-1117, 1996
A:Title: Dipeptidyl peptidase IV from Xanthomonas maltophilia: Sequencing and expression
A:Reference number: JC5142; MUID:97164011; PMID:9010758
A:Accession: JC5142
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-741 <AB>
A:Cross-references: UNIPROT:P95782; UNIPARC:UPI0000B5F8A; DDBJ:DB32263; NID:91753196; PI
C:Comment: This enzyme catalyzes the hydrolytic removal of amino-terminal dipeptidyl res
C:Superfamily: dipeptidyl-peptidase IV
C:Keywords: dipeptidylpeptide hydrolase; membrane bound
F:4-18/Domain: transmembrane #status predicted <TM>
F:610/Active site: Ser #status predicted
F:685/Active site: Asp #status predicted
F:717/Active site: His #status predicted

Query Match 17.6%; Score 708; DB 2; Length 741;
Best Local Similarity 28.3%; Pred. No. 5e-38;
Matches 208; Conservative 132; Mismatches 303; Indels 92; Gaps 28;
QY 23 TVVQKQAITPSTRSVQWLSGAEDG-----SYVYAAEDGSLTIEN-IVTNESRTILIPADK 74
Db 38 TLMKPKVADGSRVFLRGKSDRNQDLWSVDIGSGQTRLVDSQVLVPGTETLSDEEK 97
QY 75 IPTGKEAF-----NYWTHPDLSSVLMASNHTKQYRHSFFADYVQDV--BSLSKVPLM 125
Db 98 ARRRQRIAAMTGIVDYQWSDAQRLFLPLG-----GELYLDLKBGKAAVRQL 147
QY 126 PQEGDIQYAOQSPVGNITAFVRENDLYWD--NGTVTITDDGPDMEFHGPDMIYEE 183
Db 148 THGEGPATDAKLSPGGFFSFIRGNLWIDLASGRQMQLTADGSTTIGNGIAEFVADEE 207
QY 184 ILGDYR-ALWFSPDGGEYLAYSFNETGVPTY--TVQYMDNQEIAPAYPWELKIRYPKVSQ 241
Db 208 M--DRHTGTVWAPDDSAIYARIDESFVQKRYEVIADRTDV-----IEQRPAAGD 258
QY 242 TNPVTLSLNTASKEVKQAPIDAFESTDLIIGEVAWLTDTHTTVAAKAFNRVQDQK-- 299
Db 259 ANVQVKLGIVSPAEQAQATQW-IDLGKEQDIYLARVNRDPQHLSPQRSQ-----RDQKLD 313
```

```
QY 300 VVAVDTASNKATVTSDDRGTDGWLNDLSMKYIGIPKPSDKDQAYIDISHSQWHLYL 359
Db 314 LVEVTLASNOQRVLA-HETSPVTWPLHNSLRF-----DDGSILWSSERTGFQHLVRI 365
QY 360 PVSGEPIPLPKDWEVTIISIDQEROLVYVYLSLQHHSTERHLSVSVSTFAVTPLVDD 419
Db 366 D-SKGKAAALTHGNWSVDLAVDEKAGLAYFRAGIESARBSQIYAVPLQ--GGQPQRLS 422
QY 420 TVAAWMSASFANSYIILTYGGDPVYQ-ELYTTNSTKPLRTITDIAKVLQIKDYALP 478
Db 423 KAPGHSASFARNASVYVDSHNSSTFPQIELFRANGKEK-TATLVEN-----DLADPKHP 476
QY 479 NITYFELPLP-----SGET-LNVNQRLLPGGSPDKYPILFTPYGGGAGQEVTKRW 528
Db 477 YARYEARQRPVEFGTLTAADGKTPLYNSVIKPAQFPDAKRYPAVAVYVGGPASQTVTDSW 536
QY 529 QALN---FKAYVASDSLEYVTVVNRGCTGFKGRKFRSAVTROGLLEAEQIYAAQOA 585
Db 537 PGRGDLHFNQYLAQGG---YVFSLDNRGTPRRGRDFGGLYKGGKQTVVEADOLRGVWL 593
QY 586 ANIPWIDADHIGWGSFGGYLTSSKLEKDSGAPTLGVITAPVSDMRFYDSMYTERYMK 645
Db 594 KQOPWVDPARIGVQVGSNGGWTMLLAKASDSYACGVAGAPVTDWGLYDSHYTERYMDL 653
QY 646 LSTNEEGYETSAVRKTDGFKNVGEG---GFLIQHGTGDDNVHFQNSAALVDLLMGDEVSP 702
Db 654 PARNDAGYREARV-----LTHIEGLRSPLLLIHGMAADDNVLTNSTLSMALQKRG-QPF 707
QY 703 KLHSQWFTDSDHGIS 717
Db 708 ELMT--YFGAKHGLS 720

RESULT 14
T37700
probable dipeptidyl aminopeptidase - fission yeast (Schizosaccharomyces pombe) (fragment
C:Species: Schizosaccharomyces pombe
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Sep-2000
C:Accession: T37700
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1997
A:Reference number: Z21737
A:Accession: T37700
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-743 <DEV>
A:Cross-references: UNIPARC:UPI000006A67C; EMBL:Z98596; PIDN:CAB11208.1; GSPDB:GN000066;
A:Experimental source: strain 972h-; cosmid c14C4
C:Genetics:
A:Gene: SPDB:SPAC14C4.15C
A:Map position: 1

Query Match 14.5%; Score 584; DB 2; Length 743;
Best Local Similarity 23.2%; Pred. No. 5.8e-30;
Matches 177; Conservative 154; Mismatches 287; Indels 144; Gaps 26;
QY 20 FNETVVQKAITPSTRSVQWLSGAEDGSYVY-----AAEDGSLTIENIVTNE 65
Db 6 FGESI-----FLPYHODIEWITSTE-GTVLYYDQSTFSLFYPDGKEYGS-NVDSLITS- 58
QY 66 SRTLIPADKIPITGKEAFNYWIHPDLSSVLMASNHTKQYRHSFFADYVQDVBSLSKVPLM 125
Db 59 -----FVLTKNLHRKRYSSDMEYIAFSCSKRRWHSYVEDVYLVERATGRIEHLA 110
QY 126 PQEGDIQYAOQSPVGNITAFVRENDLYWDNGT--VTRITDDGGPD-MFHGVPDWIYEE 182
Db 111 SDQSKKIVVAEWSPIGHKLVLVGLGSLNLFWESFSEPPVCITDQSDLDGLFNGNSDWVYEE 170
QY 183 EILGDRYALWFSPDGGEYLAYSFNETGVPTYTVQYMDNQEIAPAYPWELKIRYPKVSQT 242
Db 171 EILQSSKAVWWSPDGNCGLSYLSIDDSKVEPHVLPEQLDLSKVEDQNRVNNFFHYSTPKDP 230
```

243 NPTVTLSLNLINIAKSEVKQAPIDA---FESTDLIIIGEVAMLTDTHTTVAAKNRV--ODQ 297
Db :
231 IPEVKL-FNCFCTDSDGEIVDSDFPLSTQHRYITDVAWAGNEKL-----NFVEVLRGNY 284
Qy :
298 QKVAVDVTASNKATVISDRDGTG--WLDNLLSMKY-----IGPIKPSDKDAYIIDISDH 350
Db :
285 ERTVSIFDLSSRKTTTENTEVSEHPALATSLLHLKXLPESLGNLKERYVRQYFL----- 339
Qy :
351 SGWAHLYLFPVSGEGEPIPLTKGDWEVTSILSIDQEOLVYYLSTQHSHTERHLYSVSYST 410
Db :
340 SNKKRIAYELDNVPFIYLTPVNISFLSDLXL--INNLYFTAISSGSPFSRYYRLCTKS 397
Qy :
411 FAVTPLVDVTVAAWYASGANSANGYILTYGGDPVPOELYTTNSTKPLRTITDNAK--- 467
Db :
398 LILLSEINIQTSGSLFGIKVNSNDQNYLLNVYLGPEIPROFYIISHBK-VSTSNHSHKNWL 456
Qy :
468 -----VLEQIK-----DYALPNITYFELPSPGETLVNMQRLLPPG 502
Db :
457 PSDSSSTSLGKVKLELCNSLETNEELIITKEKFAFFSV-FFKVIKVNITAVIQEIIRPN 515
Qy :
503 FSPDKKYPILFTPYGGGAQEVTKRWQ-----ALMFKA-----YVAS 539
Db :
516 ENPKRYPTVFHYLGAPQSALVTGKTEMDINELMASVYNFLVIKVDIVDISDVGSHLES 575
Qy :
540 DS-ELEVVTWTVDNRGTGFGRKFRSAVTRQLGLEAEDOIYAAQQAAANIPWDADHIGI 598
Db :
576 DSHELLIKSW-----IELLSRVDTPTYIDHRVGI 605
Qy :
599 WGMWSFGGLYTSKVLKEDSGAFTLGVTAPVSDWRFDVDSMYTERMYKTLSNREG-YETSA 657
Db :
606 WGMWSFGGLYTLKILE-NADPIKTAGVAPVTDWRYDAYYSENLLGAYSQKTTAIYDKTA 664
Qy :
658 VRKTDGFKNVGGGFLIOHGTCDDNVHFQNSAALVDLLMGDGV 699
Db :
665 VHYSENFRKL-CGLLVHLGTSDDNVHIENTWQLTKAMVEKGV 705

RESULT 15
T25173
hypothetical protein T23f1.7a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T25173
R:Wilkinson, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19990
A:Accession: T25173
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-779 <Wtl>
A:Cross-references: UNIPROT:O18119; UNIPARC:UIO000002A220; EMBL:Z81129; PIDN:CAB03411.1
A:Experimental source: clone T23f1
C:Genetics:
A:Gene: CESP:T23f1.7a
A:Map position: 5
A:Introns: 13/3; 52/1; 111/3; 151/2; 177/3; 346/3; 504/1; 537/1; 628/2; 603/1; 723/1
A:Superfamily: dipeptidyl-peptidase IV

	Query Match	13.5%;	Score 545;	DB 2;	Length 779;	
	Best Local Similarity	28.0%;	Pred. No. 2.le-27;			
	Matches 180;	Conservative 105;	Mismatches 245;	Indels 114;	Gaps 27;	
Qy	128	QEGDIQYAQMSPVGNNTIAFVRENDLIYV---	DNGVTVTRITDDGGPDMFHGVPDMDIYEBEI	184		
	:	:::::	:::::	:::::	:::::	
Db	161	EESVIQAQKMGKFNDFFVEVSNNKIYYOSSPEEGLTRVS--	NGGBHTVDGLFDMDIYEBEI	219		
	:	:::::	:::::	:::::	:::::	
Qy	185	LGDRAVALWFSDPGEVLAVLSFNFTGPVTTVQVYMNDQEIAPAYPWELKIRPKY--	SQTN	243		
	:	:::::	:::::	:::::	:::::	
Db	220	FGRKDAMMWSTKGDDQLAVASYNHULTXNVSLKTY----	HRLRP-YPIDTNFWHPKTFKAVL	275		
	:	:::::	:::::	:::::	:::::	
Qy	244	PTVTLSSLNLIASKEVKQAIPIDAFESTDL--II	GEVAWLTDTHTTVAAKAF-NRVDPQQKV	301		
	:	:::::	:::::	:::::	:::::	
Db	276	PTVTLSTNKTKTESSQLDVQIKDSISVHYLIAVKWLRINGTFOGVSVMTNYONEVALT	335			
	:	:::::	:::::	:::::	:::::	

```

302 QY AVD-----TASNKATVISDRDGTGMDLNLMSKYGIPKSPDKDAYIIDIS 348
      |      |      |      |      |      |      |      |      |
336 Db ICDWDTAI CRLEPEYKYASKRWTHDDPHSITSFEDTLFF-----LLPHDK----- 381
      |      |      |      |      |      |      |      |      |
349 QY DHSGWAHLIYFPVSGGE---PIPIUTKGDWEVTSILSDOEROLVYIYSTQHHSTERHLYS 405
      |      |      |      |      |      |      |      |      |
382 Db RDNAFOQVASLRLSHGQLRTPKFLNLGEYDVTSINGINKETRTIFFHAAAPKPSHRSLFS 441
      |      |      |      |      |      |      |      |      |
406 QY VSYSTFAVTPLVDDTVAAAYSASFANSGY-----YIITYGGP 443
      |      |      |      |      |      |      |      |      |
442 Db YS-----LADE-----SRNSAYCISIKNCTWAQAMDMDOMKTAIVSCKGP 483
      |      |      |      |      |      |      |      |      |
444 QY DVPYQ---ELYTTNSTKPLRTITONAKVL-----EQIKDYALPNIYTFELPLPSGETLN 494
      |      |      |      |      |      |      |      |      |
484 Db AAPHTAI VNLTRMDSDKK---TEHANLLYDKTYQNRVEEAGLPVIKETIKISDDFAL 539
      |      |      |      |      |      |      |      |      |
495 QY VMQRLPQG-FSPDK-KYPILFTEYGGPGAGEVTKRWOALNFKAYVASDSELEYVTWTVVD 551
      |      |      |      |      |      |      |      |      |
540 Db IKLSIPKDIYNRDKHQALPLIVHYVGGPDQN-TKEATQIGIEEVVASQAAIL--RID 596
      |      |      |      |      |      |      |      |      |
552 QY NRGTFGKRKPRSAVTRQLGLEAEDQIYAAQOANI--PMIDADHIGIWGWSFGYL-T 608
      |      |      |      |      |      |      |      |      |
597 Db GRGSGGRGWKYSAIYGLGTVVEBDOIKAIKVLYRLYRHLLDARXAVFGWSYGGFWTL 656
      |      |      |      |      |      |      |      |      |
609 QY SKVLEKDSGATFLGVIYIAPVSDWRFYDSMYTRYMKLTLSINEEGYETSAVRKTDFKNVE 668
      |      |      |      |      |      |      |      |      |
657 Db SMVNEAPEQFKAQVSAVPNTFNAYDAYTYTRYMGDAPL--ESY-SDVTKCLDNFKSTR 713
      |      |      |      |      |      |      |      |      |
669 QY GGFLIOHGTGDDNVHFQNSAALVDLLMGDGVS-----PEKLHS 706
      |      |      |      |      |      |      |      |      |
714 Db --LLIMHGLLDDNVHFQNSAILIDELQNRGVDFLMVTYFNQAHS 755

```

Search completed: January 9, 2006, 03:20:54
Job time : 27 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2006, 00:27:19 ; Search time 85 Seconds
(without alignments)
3902.717 Million cell updates/sec

Title: US-10-723-807-2_COPY_17_771

Perfect score: 4033

Sequence: 1 LDVPRKPHAPRGEGSKRLTF.....EKNRQTVLMHQWTKKDLLE 755

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq 21:*

1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4033	100.0	771	2	Aaw89589
2	3886	96.4	771	2	Aaw97798
3	2914	72.3	764	8	AdS23769
4	2610.5	64.7	775	9	AdY51819
5	1269	31.5	790	8	AdN21150
6	1162	28.8	793	8	AdN19590
7	1118	27.7	634	8	AdS24271
8	1054	26.1	818	8	AdS43599
9	1020	25.3	504	5	AdI17327
10	987.5	24.5	759	2	AAR54612
11	987.5	24.5	766	2	AAR40909
12	987.5	24.5	766	5	ABH08991
13	987.5	24.5	766	5	AG78417
14	987.5	24.5	766	7	ADD27855
15	987.5	24.5	766	7	ADDA6934
16	987.5	24.5	766	7	ADN39272
17	987.5	24.5	766	8	ADJ83981
18	987.5	24.5	766	8	ADJ75313
19	987.5	24.5	766	8	ADQ19398
20	987.5	24.5	766	8	ADO19806
21	987.5	24.5	766	8	ADO71612
22	987.5	24.5	766	8	ADO71644
23	987.5	24.5	766	8	ABM80355
24	987.5	24.5	766	8	ADP54458

25	987.5	24.5	766	8	ADU06688	Novel bro
26	987.5	24.5	766	8	ADV25525	Human dip
27	987.5	24.5	766	9	ADY15161	PRO polyp
28	987.5	24.5	766	9	ADY16580	PRO polyp
29	987.5	24.5	766	9	ADZ14038	Human dip
30	987.5	24.5	766	9	ABE94223	CD26/dipe
31	986.5	24.5	766	9	AA54611	Native CD
32	982	24.3	734	9	ABE94218	Human sol
33	982	24.3	760	8	ADN95552	Human BEC
34	982	24.3	760	8	ADQ21351	Human sof
35	982	24.3	760	9	ABE94159	Human wil
36	981.5	24.3	723	9	ABE94227	Human sol
37	981.5	24.3	736	8	ADO40240	Human DPP
38	981.5	24.3	766	5	ABG61910	Prostate
39	981.5	24.3	766	5	AAO15555	Human dip
40	981.5	24.3	766	6	ABP55629	Human dpp
41	981.5	24.3	766	6	ABP56700	Human liv
42	981.5	24.3	766	7	ADD14045	Human src
43	981.5	24.3	766	7	ADN39604	Cancer/an
44	981.5	24.3	766	8	ADO19400	Human PRO
45	981.5	24.3	766	8	ADQ80365	Dipeptidyl

ALIGNMENTS

RESULT 1

AAW89589
ID AAW89589 standard; protein; 771 AA.

AC AAW89589;

DT 17-MAR-1999 (first entry)

DE Aspergillus oryzae dipeptidyl aminopeptidase #1.

XX Prolyl dipeptidyl aminopeptidase; protein hydrolysate; dough;
KW flavour enhancer; palatability; mouthfeel; aroma; crust colour; baking;
KW animal feed additive; hydrolysis.

XX Aspergillus oryzae.

XX WO9851803-Al.

XX 19-NOV-1998.

XX PF 12-MAY-1998; 98WO-US009629.

XX PR 16-MAY-1997; 97US-00857884.

XX PR 20-OCT-1997; 97US-0062892P.

XX (NOVO) NOVO NORDISK BIOTECH INC.

XX Blinkovsky A, Brown K, Rey MW, Klotz A, Byun T;

XX WPI; 1999-045232/04.

XX N-PSDB; AAX00013.

XX New dipeptidyl aminopeptidase from Aspergillus oryzae - used to produce
PT protein hydrolysates enriched in particular amino acids, useful as
PT flavour enhancers, e.g. in doughs.

XX Claim 1; Page 56-57; 77pp; English.

XX The present sequence represents dipeptidyl aminopeptidase (DPAP) from
CC Aspergillus oryzae. DPAP acts synergistically with an aminopeptidase (AP)
CC to hydrolyse polypeptides, producing protein hydrolysate (pH), useful in
CC foods as flavour enhancer, e.g. in baked goods, enriched in: (a) Ala,
CC Arg Asp, Gly and/or Val, or (b), if the substrate has been deamidated,
CC in Glu (free and/or peptide bound), in which case products are useful as
CC animal feed additives. DPAP can also be used in flavour- improving
CC compositions (optionally containing AP) and in dough pre-mixes, also for
CC deactivating enzymes and for converting precursors to mature proteins.

CC DPAP increases the level of hydrolysis of proteins and thus of flavour
CC development, and a mixture with AP may hydrolyse tripeptides that are
CC resistant to either enzyme used alone. PH have improved solubility,
CC emulsifying and foaming properties, and products containing them have
CC better flavour, palatability, mouthfeel, aroma and crust colour
XX
SQ Sequence 771 AA;

Query Match 100.0%; Score 4033; DB 2; Length 771;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LDVPRKPHAPTGEKSKLTFNETVVKQAITPTSRVQWLSGAEDGSVYAAEDGSLTIEN 60
Db 17 LDVPRKPHAPTGEKSKLTFNETVVKQAITPTSRVQWLSGAEDGSVYAAEDGSLTIEN 76
Qy 61 IVTNESRTLPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYVVDVESLK 120
Db 77 IVTNESRTLPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYVVDVESLK 136
Qy 121 SVPLMPDQEGDIQYAAQWSPVGNNTIAFVRENDLYVMDNGTVTRITDDGGPMFHGVPDWIY 180
Db 137 SVPLMPDQEGDIQYAAQWSPVGNNTIAFVRENDLYVMDNGTVTRITDDGGPMFHGVPDWIY 196
Qy 181 EEEILGDRYALWFSPDGEYLAYLSFNETGVPTTYVQYMDNQETAPAYPWEKIRYPKVS 240
Db 197 EEEILGDRYALWFSPDGEYLAYLSFNETGVPTTYVQYMDNQETAPAYPWEKIRYPKVS 256
Qy 241 QTNPTVTLSLNIASKEVKQAPIDAFESTDLICEVAWLTDTHTTVAAKAFNRVQDOQKV 300
Db 257 QTNPTVTLSLNIASKEVKQAPIDAFESTDLICEVAWLTDTHTTVAAKAFNRVQDOQKV 316
Qy 301 VAVDTASNKATVSDRGTGDLNLLSMKYIGIPKSPDKDAYYIDISDHSGWAHLYLFP 360
Db 317 VAVDTASNKATVSDRGTGDLNLLSMKYIGIPKSPDKDAYYIDISDHSGWAHLYLFP 376
Qy 361 VSGGEPIPLTKGDWEVTSILSIDERQLVYVLSLTHQHSSTERHLYSVSYSTFAVTPLVDDT 420
Db 377 VSGGEPIPLTKGDWEVTSILSIDERQLVYVLSLTHQHSSTERHLYSVSYSTFAVTPLVDDT 436
Qy 421 VAAWKSASFANSQVYIITYGGPDVPVQELVTTNSTKPLRTITONAKVLEQIKDYALPNI 480
Db 437 VAAWKSASFANSQVYIITYGGPDVPVQELVTTNSTKPLRTITONAKVLEQIKDYALPNI 496
Qy 481 TYFELPLPSGETLNMQRLLPFGSPDKKYPILFTPYGPGQAQEVTKRWQALNFKAYVASD 540
Db 497 TYFELPLPSGETLNMQRLLPFGSPDKKYPILFTPYGPGQAQEVTKRWQALNFKAYVASD 556
Qy 541 SELEYVTWVDRNGTGPKGRFSAVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGIWG 600
Db 557 SELEYVTWVDRNGTGPKGRFSAVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGIWG 616
Qy 601 WSPFGYLTSKVLKXDSGAFITGVITAPVSDWRFRVDSMYTERYMKTLSTNBEGETSAVRK 660
Db 617 WSPFGYLTSKVLKXDSGAFITGVITAPVSDWRFRVDSMYTERYMKTLSTNBEGETSAVRK 676
Qy 661 TDGFKNVGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSRHGISYHG 720
Db 677 TDGFKNVGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSRHGISYHG 736
Qy 721 GGVELYKQALARKLYOEKNRQTOVLMLHQWTKDLEE 755
Db 737 GGVELYKQALARKLYOEKNRQTOVLMLHQWTKDLEE 771

RESULT 2

AAW97798

ID AAW97798 standard; protein; 771 AA.

XX

AC AAW97798;

XX

DT 21-MAY-1999 (first entry)

XX

DE Aspergillus oryzae prolyl-dipeptidyl-peptidase DPPIV.
XX Prolyl-dipeptidyl-peptidase; DPPIV; protein hydrolysate; food.
XX Aspergillus oryzae.
XX
PH Key Location/Qualifiers
FT Peptide 1..16
FT Protein /note= "signal peptide, specifically claimed in Claim 4"
FT /note= "mature protein, specifically claimed in Claim 1"
XX
XX WO9902705-A1.
XX 21-JAN-1999.
XX 06-MAY-1998; 98WO-EP002799.
XX 05-JUL-1997; 97EP-00111377.
XX (NEST) SOC PROD NESTLE SA.
XX Monod M, Doumas A, Affolter M, Van Den Broek P;
XX WPI; 1999-120906/10.
XX N-PSDB; AAX07327.
XX New Aspergillus oryzae prolyl-dipeptidyl-peptidase enzyme (DPPIV) -
XX having a high level of hydrolysing specificity towards proteins starting
XX with X-Pro, where X is any amino acid.
XX Claim 1; Page 27-29; 41pp; English.
XX This is the amino acid sequence of the prolyl-dipeptidyl-peptidase
XX (DPPIV) of Aspergillus niger strain 44. DPPIV is a secreted enzyme having
XX a high level of hydrolysing specificity towards proteins starting with X-
XX Pro, where X is any amino acid. The DPPIV gene (see AAX07327) was
XX isolated by screening a genomic DNA library with a DNA fragment containing
XX the DPPIV gene of Aspergillus fumigatus. The invention also provides host
XX cells able to over-express the enzyme, providing at least 50 mU of DPPIV
XX activity/ml of supernatant when grown in minimal medium containing 1
XX wt./vol% of wheat gluten, as well as Aspergillus cells having integrated
XX multiple copies of the DPPIV gene promoter, recombinant prolyl-dipeptidyl
XX -peptidase having the leader sequence of DPPIV, and a method for
XX producing the enzyme by cultivating recombinant cells. The DPPIV enzyme
XX or cells expressing it can be used, optionally in conjunction with
XX another proliadase, to hydrolyse protein containing materials. Food
XX products can be a milk substitute for infants, a hydrolysed vegetable
XX protein (koji), a leguminous or oleaginous plant, or cooked or roasted
XX cereal sources. If the DPPIV activity is combined with other proteolytic
XX activities, a high degree of hydrolysis may be obtained leading to a non-
XX bitter flavour and a significantly lower allergenicity than unhydrolysed
XX proteins
SQ Sequence 771 AA;

Query Match 96.4%; Score 3886; DB 2; Length 771;
Best Local Similarity 96.8%; Pred. No. 6.1e-311;
Matches 733; Conservative 1; Mismatches 19; Indels 4; Gaps 2;

Qy 1 LDVPRKPHAPTGEKSKLTFNETVVKQAITPTSRVQWLSGAEDGSVYAAEDGSLTIEN 60
Db 17 LDVPRKPHAPTGEKSKLTFNETVVKQAITPTSRVQWLSGAEDGS--LRVGRGRRSHH 74
Qy 61 IVTNESRT--LIPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYVVDVES 118
Db 75 REHHQRVTHAHPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYVVDVES 134
Qy 119 LKSVPLMPDQEGDIQYAAQWSPVGNNTIAFVRENDLYVMDNGTVTRITDDGGPMFHGVPDW 178
Db 135 LKSVPLMPDQEGDIQYAAQWSPVGNNTIAFVRENDLYVMDNGTVTRITDDGGPMFHGVPDW 194
Qy 179 IYBEEILGDRYALWFSPDGEYLAYLSFNETGVPTTYVQYMDNQEIAPAYPWEKIRYPK 238

```
Db 195 IYEEILGDRYALWSPDGEYLAISFNETGVPTVTVQYMDNQEIADAPWELKIRYPK 254
Qy 239 VSQTNPTVTLNLNLTASKEVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDQ 298
Db 255 VSQTNPTVTLNLNLTASKEVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDQ 314
Qy 299 KVAVDVTASNKATVLSRDRDGTGMDLNLNLSMKYIGPIKPSDKDAYIIDSJHSGWAHLYL 358
Db 315 KVAVDVTASNKATVLSRDRDGTGMDLNLNLSMKYIGPIKPSDKDAYIIDSJHSGWAHLYL 374
Qy 359 FVPSGGEPIPLTKGDWEVTSILSDQERQLVYLSSTQHSTHERHLYSVSYSTFAVTPLVD 418
Db 375 FVPSGGEPIPLTKGDWEVTSILSDQERQLVYLSSTQHSTHERHLYSVSYSTFAVTPLVD 434
Qy 419 DTVAAYWASFSANSFYILTYGGPDVYQELTYTNSKPLRTITDNAKVLQIKDYALP 478
Db 435 DTVAAYWASFSANSFYILTYGGPDVYQELTYTNSKPLRTITDNAKVLQIKDYALP 494
Qy 479 NITYFELPLPSGETLNVMQRLPPGSPDKKYPIILFTPYGGCAQEVTKRWQALNFKAYVA 538
Db 495 NITYFELPLPSGETLNVMQRLPPGSPDKKYPIILFTPYGGCAQEVTKRWQALNFKAYVA 554
Qy 539 SDSELEYVTWVDNRGTGPKGRKFRSAVTRQLGLEAEQIYAAQQAANIPIWDADHIGI 598
Db 555 SDSELEYVTWVDNRGTGPKGRKFRSAVTRQLGLEAEQIYAAQQAANIPIWDADHIGI 614
Qy 599 WQWSEGGYLTSTKLEKSDGAFITGVITAPVSDWRFVDSMYTRYMYTKLSTNEEGYETSAV 658
Db 615 WQWSEGGYLTSTKLEKSDGAFITGVITAPVSDWRFVDSMYTRYMYTKLSTNEEGYETSAV 674
Qy 659 RTIDGFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSQWFTDSDHGISY 718
Db 675 RTIDGFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSQWFTDSDHGISY 734
Qy 719 HGGGVFLYKQARKLYQEKNRQTQVLMHQWTKKDL 755
Db 735 HGGGVFLYKQARKLYQEKNRQTQVLMHQWTKKDL 771

RESULT 3
ADS23769
ID ADS23769 standard; protein; 764 AA.
XX ADS23769;
AC ADS23769;
DT 02-DEC-2004 (first entry)
XX Bacterial polypeptide #12802.
DE
XX Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
XX US2003233675-A1.
XX
XX 18-DEC-2003.
XX
XX 20-FEB-2003; 2003US-00369493.
XX
XX 21-FEB-2002; 2002US-0360039P.
XX
XX (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
XX (SLAT/) SLATER S C.
XX (CHEN/) CHEN X.
XX (GOLD/) GOLDMAN B S.
```

```
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
XX for expression of a polynucleotide encoding a polypeptide from a
XX microbial source, useful for producing plants with improved properties.
XX
XX Claim 1; SEQ ID NO 12802; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
XX promoter functional in a plant cell, where the promoter is positioned to
XX provide for expression of a polynucleotide encoding a polypeptide from a
XX microbial source. The invention also relates to a transformed plant
XX comprising the recombinant DNA construct and a method of producing a
XX transformed plant having an improved property. The plant is a crop plant
XX such as maize or soybean. The method of producing a transformed plant
XX having an improved property comprises transforming a plant with the
XX recombinant DNA construct and growing the transformed plant, where the
XX polynucleotide or polypeptide is useful for improving plant properties.
XX The recombinant DNA construct is useful for producing plants with
XX improved plant properties, e.g. improved cold, heat or drought tolerance,
XX tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX increased resistance to plant disease, better growth rate by modification
XX of the cell cycle pathway with plant growth regulators, increased rate of
XX homologous recombination, modified seed oil or protein yield and/or
XX content, improved yield by modification of carbohydrate, nitrogen or by
XX phosphorus use and/or uptake, by modification of photosynthesis or by
XX providing improved plant growth and development under at least one stress
XX condition, improved lignin production or improved galactomannan
XX production. This sequence represents a bacterial polypeptide used in the
XX scope of the invention. Note: The sequence data for this patent did not
XX form part of the printed specification but was obtained in electronic
XX format from USPTO at seqdata.uspto.gov/sequence.html.
XX
XX Sequence 764 AA;
XX
XX Query Match 72.3%; Score 2914; DB 8; Length 764;
XX Best Local Similarity 69.3%; Pred. No. 9.2e-231;
XX Matches 524; Conservative 103; Mismatches 113; Indels 16; Gaps 4;
XX
Qy 4 PRKHAPTEGSGKRLTFNETVVKQAITPTSRVSQVLSGAEKDSYV-----YAAEDGSLTI 58
Db 17 PWRPREPRAAGSKRLTFNETVISAALSPPSSISVQVIATENDGGCLCVPGGKWHQDKEHCH 76
Qy 59 ENIVTNESERTLIPADKIPGKEAFNYWHPDLSVLSWASNHTKQYRHSFFADYVQDVES 118
Db 77 QPIADNRSABKIPAD-----AYSTWISPDLSAVLMATNYTKQYRHSFFADYVQDVET 129
Qy 119 LKSVPLMPDQEGDIQYAWSPVGNITAFVRENDLYVNDNGTVTRITDDGDPDMFHGVPDW 178
Db 130 LFTVPLVEDMVGDIQYAEWSPSGDSIAFVRGNLMTWSDGTVTATIKDGGPDMFHGVPDW 189
Qy 179 IYEEILGDRYALWSPDGEYLAISFNETGVPTVTVQYMDNQEIADAPWELKIRYPK 238
Db 190 IYEEILGDRYALWSPDSELLAFLTFNETGVPTFTVQYFMDNQEIADAPWELKIRYPK 249
Qy 239 VSQTNPTVTLNLNLTASKEVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDQ 298
Db 250 VSETNPTVKLNILQLQSDNTVSTIPDIVDFPSELIVGEVAWVTDTHTELAVKAFNRVQDES 309
Qy 299 KVAVDVTASNKATVLSRDRDGTGMDLNLNLSMKYIGPIKPSDKD---AYYIDISDHSGWAH 355
Db 310 KVIIVETASGETKIAHERDGTGMDLNLNLSISVGFALGSGDASSAYYVLDSDHSGWTH 369
Qy 356 LYLFPVSGGEPIPLTKGDWEVTSILSDQERQLVYLSSTQHSTHERHLYSVSYSTFAVT 415
Db 370 LYLFPSTSGDPIPLTEGEWEVTSIVSDQERELFYLSSTQHSTHERHLYSVSYRTFEITP 429
Qy 416 LVDDTVAAYWASFSANSFYILTYGGPDVYQELTYTNSKPLRTITDNAKVLQIKDY 475
Db 430 LVDDTVEAYWSVFSKAGYYILTYAGPSVPYQELYSVNVQATPLRTILTSNAILIEKLE 489
```


KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.

OS Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINK/) HINKLE G J.

XX (SLAT/) SLATER S C.

XX (CHEN/) CHEN X.

XX (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 3803; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polypeptide used in the
 CC scope of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 790 AA;

Query Match 31.5%; Score 1269; DB 8; Length 790;

Best Local Similarity 38.7%; Pred. No. 4.9e-95;

Matches 296; Conservative 132; Mismatches 271; Indels 66; Gaps 29;

QY 7 PHAPTGEKS-KRLTFNETVVKQAITPTSRVQWLSG--AEDG-----SYVYAAED 53

DB 34 PQATVTHGSGKVTLDQ-VLNNWRKASISWAGNAGEDGLLLEKEGANKDYLVVEDV 92

QY 54 GSLTNIENVNSTRILIPADKIPTGKEAFN--YWIHP-----DLSSVLWASNHTKQVR 104

DB 93 RAQNPSSVEASKSTLI-KDKL---FEFANKTYW--PTVTPSRDLKKVLLATDVQNNWR 146

QY 105 HSPFADYVVDVSLKSVFLMP--DQEGDIQYAKWSPVGNVIAFVRENDLV--WDNGTVT 161

DB 147 HSYAVVWTFDVTQAEPLVPYDADARLQLASWSPSDAIVVTRDNMFLRLKSDKIV 206

QY 162 RITDDGDPNFHGVDPDIYEEILGDRYALWFSFDGDEYLAISFNETGVPTTVQVYM-- 219

DB 207 QITRDGSADVFNQVDPDWMYEEVLASGVATWMSDGNVAFURTNETGVPEPIQYFVSR 266

QY 220 -DNQEIAPA---YPWELKIRYPKVSQTNPTVTLSELNIAASKVKQAPIDA-FESTDLIIG 274

DB 267 PSGEFPKPEENYPEVRQIKTPKAGAHNPVDLKFYDVKRGDVFSVDISGRADDDRLIT 326

QY 275 EVANLTDTHTTVAAKAFNRVQDQKQVAVDTASN-----KATVISDRDGTGMLNLLSM 329

DB 327 EVIM---AGQVLIKETNRVSDVMRVLVDVGSRTGKAVRTVDVNDIDG--GWFEISHKT 381

QY 330 KYIGPIKPSD--KDAYYIDIDSHSWAHL-YLFPVSGGEPIPLTKGDWEVTIILS-IDOE 385

DB 382 KFI-PADPANGRPDDGYVDTIHHNGDHLAYFTPLDNPNPIMLTSGDYEVVDAPSVDIQ 440

QY 386 ROLVYVLTSTQHHSTERHLYSVSYSTFAVTPLVDDTVAAVWSASFANSYGYILTYGPDV 445

DB 441 RNLVYFVSTKESSIORHYQVKLTGEDMTPTVDTSKGYVYALSFSTGAGYALVSYQGPNI 500

QY 446 PYQELTYT--NSTKPLRTITDNAKVLEQIKOYALPNITYFELPLPSGETLNVMQRLPPGF 503

DB 501 PMQKVIPTSPNDKYEHVVEENKDLAAKAKHELPIINYGTINV-DGVELNVIERRPPHF 559

QY 504 SPDKKYPIILFTPYGPGAGQEVTKRQALNFKAYVASDSSELYVTVTVNDRGTGFKGRKR 563

DB 560 DNKKKYPVLFOQYSGFVSQTVKTF-AVDQSFVAAG--LGVICVTVDRGTGFGIGRKNR 616

QY 564 SAVTRQLGLEAEDQIYAAQQAANIPWIDADHIGTWGVSFGVLTSTKVLKDSG-AFTLG 622

DB 617 VTIRNGLTWESHDQIAAKHWAQKDYIDEDRLATWGSYGGYMTLKTLEQDAGQTFKYG 676

QY 623 VITAPVSDWRFVDSMYTERYMTKLTSTNEEGYETSARVKTDPG-KNVEGGFLIOHGTGDDN 691

DB 677 MAVAPVTDWRFVDSIYTERYMETPQTPNEGYSAAVTNVLTALSQNVN--FLLMHGVADDN 734

QY 682 VHFQNSAALVDLLMGDGVSPKHLHSONFTDSDHSGISYHGGGVFLY 726

DB 735 VHMQNSLTLLDAL--DQRSVENYDVQVPDSDHGIYFHNANRIVF 777

RESULT 6

ADN19590

ID ADN19590 standard; protein; 793 AA.

XX AC ADN19590;

XX DT 02-DEC-2004 (first entry)

XX DE Bacterial polypeptide #2243.

XX KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.

XX OS Bacteria.

XX US2003233675-A1.

XX PD 18-DEC-2003.

XX PF 20-FEB-2003; 2003US-00369493.

XX PR 21-FEB-2002; 2002US-0360039P.

XX PA (CAOY/) CAO Y.

XX PA (HINK/) HINKLE G J.

XX PA (SLAT/) SLATER S C.

XX PA (CHEN/) CHEN X.

XX PA (GOLD/) GOLDMAN B S.

pharmacogenomics.

PT Disclosure; SEQ ID NO 863; 1498pp; English.

This invention relates to a novel nucleic acids, and encoded polypeptides thereof, which have properties related to the stimulation of biochemical or physiological responses in a cell, tissue, organ or organism. Specifically, it refers to the use of biologically active fragments for diagnostic and prognostic assays and furthermore in the treatment of diverse pathological conditions. The present invention describes novel human and murine NOVX proteins, as well as methods to modulate their expression using antisense oligos, ribozymes and peptide nucleic acids. The NOVX polypeptides, polynucleotides and antibodies are useful in treating or preventing NOVX-associated disorders, e.g. cardiomyopathy, atherosclerosis, cancer and diabetes. Furthermore, they may be used in treating or preventing diseases such as inflammation, autoimmune disorders, allergies, blood disorders, acquired immunodeficiency syndrome (AIDS), obesity, asthma, immunoglobulin (Ig)A nephropathy, cirrhosis, arthritis, Alzheimer's disease, infections, stroke, muscular dystrophy and epilepsy. Accordingly, these molecules have many activities including cytostatic, cardiac, antiinflammatory, immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic, antiarteriosclerotic, anorectic, antasthmatic, nephrotropic, antiarthritic, hepatotropic, neuroprotective, nootropic, antibacterial, virucide, antiparasitic, relaxant and anticonvulsant. In addition, they are useful in screening assays to identify small molecules that modulate or inhibit, for example, neurogenesis, wound healing and angiogenesis. The nucleic acids are also used as in chromosome mapping, tissue typing, preventive medicine and pharmacogenomics. This polypeptide is a protein fragment that is homologous to a domain of a human NOVX protein of the invention.

XX Sequence 504 AA;

Query Match 25.3%; Score 1020; DB 5; Length 504;

Best Local Similarity 41.2%; Pred. No. 8.3e-75;

Matches 212; Conservative 100; Mismatches 180; Indels 22; Gaps 12;

QY 22 ETWVQAIPTRSVMWLSGAEVGSVYAAEDGSLTIENIVNESRTLIPADKIPGKEA 81
DB 4 EDIFSFTPKYDIINWIS---DGEYLQDDTNLLVYVETGKT-TVLLSDTFNEFEA 59
QY 82 FNYWIHPDLSSVLWASNHTKQYRHSFFADYYVQDVESLSKSVLPMPOQEGD-IQYAWQSPV 140
DB 60 SNVILSPDGKYLITSTYKRWHSYASYYIDLTGDPVPAPEESGKIQYATWSPK 119
QY 141 GNTIAFVRENDLVV--WDNGTVTRITDDG-GPDMFHGVDPDWYEEBILGDRYALWSPDG 197
DB 120 GHKLAFVRDNLVYIQKLPSGPAIQITTDGKSNDFNGIPDWYEEBILSTDYALWSPDG 179
QY 198 EYLAYLSFNETGYPTTVQYVMDNQEIAIPAPWELKIRYKVSQTWPTVTLNLNTA-SK 256
DB 180 DFLAYLRFNDSEVPVTEYPPYTDQSQ----YPSDMEIKYPAQDPNPTVKLFPVNLADGA 235
QY 257 EVKQAPIDA-FESTDLIIGEVAMLTDTHTTVAKAFNRVQDQKVVAVDPTASNKATVISD 315
DB 236 SVSEIPLPASLASGDIYITVAVTNER--LAVQWLNRDQNI SVLSICDTASSTWNVKN 293
QY 316 -RDGTGWLNDLSMKYIGIPKSPDKDAYIDISDHSGWAHLVFPVSGEPTPLTKGDW 374
DB 294 FEDSETGWVETFPNLSL---PVFPLDGLSYLDISDRDGYKHLAYEYLDGKEPTALTKGNW 350
QY 375 EVTSILSISOERQVLYSLTQHSHTERHLYSVS--STFAVTPLVDDTVAAYWSASFSA 432
DB 351 EVTNILGVDSKTDVTFATPEESGERHLYSLSLKGKTTLSQDLSERCGYYSASFSPN 410
QY 433 SGYYILTYGDPVPYQBYLTNTSKPLRTITDNAKLVLEQIKOYALPNITTYFELPLPSGET 492
DB 411 AKYIILTYSGFVPIQTLHSSNDTKELRTLEDNEALKKALKNYQLPSKKEGKIKLADGIT 470
QY 493 LNVMQRLPPGFSPPDKYKYPILFTPYGGGGAQEVTK 526
DB 471 LNYQMINKPANFDFSKYPVLFVYGGGSGQVTK 504

RESULT 10

AARS4612

XX AARS4612 standard; protein; 759 AA.

AC AARS4612;

DT 25-MAR-2003 (revised)

DT 09-DEC-1994 (first entry)

DE Delta3-9 CD26.

XX Human; T cell activation antigen; CD26; analogues; deletion; soluble;
KW signal peptidase; immune-stimulating; response-stimulating; AIDS;
KW immunosuppression; AIDS-related complex.

OS Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 2..3

/note= "Position of delta3-9 deletion"

XX WO9409132-A1.

PD 28-APR-1994.

XX 19-AUG-1993; 93WO-US007923.

XX 21-AUG-1992; 92US-00934162.

XX (DAND) DANA FARBER CANCER INST INC.

XX Morimoto C. Schlossman S, Tanaka T;

XX WPI; 1994-151317/18.

XX Polypeptide fragments and analogues of CD26 and encoding nucleic acid -
PT useful for stimulating immune response, e.g. for treatment of AIDS to
XX counteract immunosuppressive drug, and as vaccine adjuvant.

XX Claim 3; Page 49-52; 85pp; English.

XX The sequences given in AARS4612-14 represents analogues of the human T
cell activation antigen CD26 which have internal deletions. The analogues
pref. lack residues 3-9 or 24-34. These analogues are soluble under
physiological conditions and lack enough amino acid residues to render
them susceptible to cleavage by signal peptidase. The peptide fragments
and analogues are useful as immune or response- stimulating therapeutics,
eg. they may be used for treatment of disease conditions characterised by
immunosuppression, eg. AIDS or AIDS-related complex, other virally or
environmentally-induced conditions, and certain congenital immune
deficiencies. The peptides can be employed to increase immune function
which has been impaired by use of immunosuppressive drugs, such as certain
chemotherapeutic drugs. (Updated on 25-MAR-2003 to correct PN field.)

XX Sequence 759 AA;

Query Match 24.5%; Score 987.5; DB 2; Length 759;

Best Local Similarity 32.6%; Pred. No. 7.6e-72;

Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;

QY 11 TGEGRSLTFNFTVVKQAITPTSRVQWLSGAEVGSVYAAEDGSLTIENIVNESRTLI 70
DB 29 TADSRKTYTLTD-YLKNTRYLKLISLRWIS--DHEVLYKQENILVFAEYGNSSVFLE 84
QY 71 PADKIPTGKEAFNFWTHPDLSVLWASNHTKQYRHSFFADYYVQDVESLSKSVLPMPOE- 129
DB 85 NSTFDFGHSINDYSISPDQGFILLEYNYVVKWRHSYTASY--DIYDLNKRQLITEERI 141
QY 130 -GDIOYQAQSPGCVNTAFVRENDLVV--WDNGTVTRITDDGGPD-MFHGVDPDWYEEBIL 185
DB 142 PNTQWTVSPVGHKLAYVWNNDIYVKEIPNLPSYRITWTGKEDIYNGITDWYEEBVF 201

QY 512 LFTPYGGCAQEVTRKQALNFKAYVASDSELEYVTVTDNRGTGFKGRKFRSAVTRQLG 571
DB 543 LLDVYAGPCSQKADTVFR-LNWTYLASTENI--IVASFDGRSGYQGDKIMHAINRRLG 599
QY 572 LLEAEDQIYAAQQAANIPWIDADHIGIWMGSPGGYLTYSKVLKDSGAFTLGVTAPVSDW 631
DB 600 TFEVEDQIEAARQFSKMGFVDNKRITAIWGSYGGYVTSVVLGSGSGVFKCGIAVAPVSRW 659
QY 632 RYDMSMYTERYM--KTLSTNEEGYETSAV-RKTDGFKNVEGFLIOHGTGDDNVHFQNSA 688
DB 660 EYDVSVYTERYMGLEPTPDNLDRYNSVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 717
QY 689 ALVDLLMGDVSPEKLHSQWFTSDHGISYHGGVFLYKQLARKLYQ 735
DB 718 QISKALVDVGVDD---FQAMWYTDDEHGIASSTAHOHIYTHMSHFQ 761

RESULT 12

ABB08991

ID ABB08991 standard; protein; 766 AA.

XX

AC ABB08991;

XX

DT 19-JUN-2002 (first entry)

XX

DE Human dipeptidyl peptidase IV.

XX

DE Human; dipeptidyl peptidase IV; antiasthmatic; anti-allergic;

KW anti-inflammatory.

XX

OS Homo sapiens.

XX

PN US6337069-B1.

XX

PD 08-JAN-2002.

XX

PF 28-FEB-2001; 2001US-00794236.

XX

PR 28-FEB-2001; 2001US-00794236.

XX

PA (BMRA-) BMRA CORP BV.

XX

PI Grouzmann E, Lacroix J, Monod M;

XX

DR WPI; 2002-163235/21.

XX

PT Treating a patient for mucosal inflammation associated with rhinitis,

PT

PT sinusitis or both, by intranasally administering a peptidase that cleaves

PT

PT at Xaa-Pro sequences, to the patient.

XX

PS Disclosure; Col 9-14; 13pp; English.

XX

CC Thus invention relates to the treating of a patient for mucosal

XX

CC inflammation associated with rhinitis or sinusitis, comprising

CC

CC intranasally administering a peptidase. The peptidase is considered

CC

CC antiasthmatic, anti-allergic and anti-inflammatory in its action. The

CC

CC peptidase cleaves at Xaa-Pro sequences and is useful for treating a

CC

CC patient for mucosal inflammation associated with rhinitis or sinusitis,

CC

CC which is the result of allergies or asthma. This sequence represents

CC

CC human dipeptidyl peptidase IV

XX

SQ Sequence 766 AA;

DB 92 NSTDEFHGSINDYSISPDGQFILLEYNVYQWRRHSYTSY---DIYDLNKGQLTEERI 148
QY 130 -GDIOYAWQSPVGNITAFVRENDLYV--WDNGTIVTRITDDGGPD-MFHGVPDVIYEEEL 185
DB 149 PNNTQWVTSVPGHKLAVVWNDIYKLEPNLPSYRIITWTKGDIINGITDWMVEEVF 208
QY 186 GDRYALWFSPOGEYLAYSFNFTVPTVTVQVYMDNQEIAPAYPWLKIRYKPVSGTNT 245
DB 209 SAYSLAWSPNGTFLAYAQFNDTEVPLIEYSFYSDE---SLQYPKTVRPYKAGAVNFT 265
QY 246 VTLSLINT--ASKEVYQAPIDAFESTDLIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297
DB 266 VKFFVYVNTDSSSVTNATSIQITAPASMLIGHYICDVTWATQER--ISLQWLRRLQNY 323
QY 298 -QKVAVVATASNKATVISDRD---GTDGWLNDLLSMKYIGPIKFS-----DKDAYYTD 346
DB 324 VMDICDYDESSGRWNCVVARQHIEMSTTCW-----VGRFSEPHFTLDGNSFYKI 374
QY 347 ISHSGWAHLYLFPVSGGEPIPLTKGDWEVTSILSIDQERQLVYLLSTOHHSTE--RHLY 404
DB 375 ISNEEGYRHICVFQIDKKDCTFTTKGTWEVIGIEALTSD--YLYIISNEYKMGPGGRNLY 432
QY 405 SVSYSTFA-VTEL---VDDTVAAYHSASFSANSYVILTYGGPDVYQELTYTNSKPLR 460
DB 433 KIQLSDYTKVTCLSCELNPERCQYYSVSFSKAKYQLRCSGFGGLPYTLHSSVNDKGLR 492
QY 461 TITDNA---KVLQIK-----DYALPNITYFELPLPSGETLNMVQRLPPGSPDKKYPI 511
DB 493 VLEDNSALDKMLQNVQMPSKLDIFILNETKFWYQM-----ILPFPDKSKKYP 542
QY 512 LFTPYGGCAQEVTRKQALNFKAYVASDSELEYVTVTDNRGTGFKGRKFRSAVTRQLG 571
DB 543 LLDVYAGPCSQKADTVFR-LNWTYLASTENI--IVASFDGRSGYQGDKIMHAINRRLG 599
QY 572 LLEAEDQIYAAQQAANIPWIDADHIGIWMGSPGGYLTYSKVLKDSGAFTLGVTAPVSDW 631
DB 600 TFEVEDQIEAARQFSKMGFVDNKRITAIWGSYGGYVTSVVLGSGSGVFKCGIAVAPVSRW 659
QY 632 RYDMSMYTERYM--KTLSTNEEGYETSAV-RKTDGFKNVEGFLIOHGTGDDNVHFQNSA 688
DB 660 EYDVSVYTERYMGLEPTPDNLDRYNSVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 717
QY 689 ALVDLLMGDVSPEKLHSQWFTSDHGISYHGGVFLYKQLARKLYQ 735
DB 718 QISKALVDVGVDD---FQAMWYTDDEHGIASSTAHOHIYTHMSHFQ 761

RESULT 13

AAG78417

ID AAG78417 standard; protein; 766 AA.

XX

AC AAG78417;

XX

DT 12-APR-2002 (first entry)

XX

DE Human dipeptidyl peptidase IV amino acid sequence.

XX

XX 21953 prolly oligopeptidase; antibody; proline; endopeptidase; cancer;

KW cardiovascular disease; autoimmune disease; atopic allergy;

KW neuronal disorder; vascular disorder; prostate disorder; cytostatic;

KW antidiabetic; antiarthritic; antiasthmatic; anti-inflammatory;

KW diabetes mellitus; arthritis; multiple sclerosis; asthma;

KW Grave's disease; neuronal disorder; demyelinating disease;

KW dipeptidyl peptidase.

XX

XX Homo sapiens.

XX OS

XX WO200179473-A2.

XX

XX 25-OCT-2001.

XX

XX 11-APR-2001; 2001WO-US040483.

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2006, 02:48:49 ; Search time 90 seconds
(without alignments)
5918.599 Million cell updates/sec

Title: US-10-723-807-2_COPY_17_771

Perfect score: 4033

Sequence: 1 LDVPRKPHAPTGGSKRLTF.....EKNRQTQVLHQTWKDLEE 755

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3886	96.4	771	042812 ASPOR	042812 aspergillus
2	3020	74.9	773	05A242 EMENI	05A242 aspergillus
3	3010.5	74.6	765	04WRH9 ASPFU	04WRH9 aspergillus
4	3008.5	74.6	765	014425 ASPFU	014425 aspergillus
5	2610.5	64.7	775	05J6J3 TRIRU	05J6J3 trichophyto
6	2411.5	59.8	778	0512P7 MAGGR	0512P7 magnaporthe
7	2399.5	59.5	779	04HVM6 GIBZE	04HVM6 gibberella
8	1376	34.1	911	04IAY7 GIBZE	04IAY7 gibberella
9	1295	32.1	901	096VT7 ASPNG	096VT7 aspergillus
10	1289	32.0	874	07S180 EMENI	07S180 emericeila
11	1284	31.8	880	05B934 EMENI	05B934 aspergillus
12	1276.5	31.7	895	07SHUB NEUCR	07SHUB neurospora
13	1262.5	31.3	947	04WX13 ASPFU	04WX13 aspergillus
14	1248.5	31.0	938	051X20 MAGGR	051X20 magnaporthe
15	1225.5	30.4	868	06CFX2 YARLI	06CFX2 yarrowia li
16	1193.5	29.6	883	055H13 CRYNE	055H13 cryptococcu
17	1193.5	29.6	883	05K721 CRYNE	05K721 cryptococcu
18	1162	28.8	793	1 YEAB SCHPO	1 schizosacch
19	1155.5	28.7	870	06BP08 DEBHA	06BP08 debaryomyce
20	1150.5	28.5	829	06CMM1 KLULA	06CMM1 kluyveromyc
21	1109	27.5	817	06FPV7 CANGA	06FPV7 candida gla
22	1098	27.2	838	05AG57 CANAL	05AG57 candida alb
23	1098	27.2	838	05AGJ4 CANAL	05AGJ4 candida alb
24	1069.5	26.5	737	0641D6 XENLA	0641D6 xenopus lae
25	1062.5	26.3	872	075C44 ASHGO	075C44 ashbya goss
26	1061	26.3	748	2 P70092 XENLA	2 xenopus lae
27	1055	26.2	818	2 Q66R87 YEAST	2 saccharomyc
28	1054	26.1	818	1 DAP2 YEAST	1 P18962 saccharomyc
29	1010	25.0	759	05Z181 CHICK	05Z181 gallus gall
30	1008.5	25.0	751	2 Q75883 AGKHB	2 agkistrodon
31	1007.5	25.0	751	2 Q75882 AGKHB	2 agkistrodon

32	991	24.6	766	1 DPP4_PIG	P22411 s dipeptidy
33	987.5	24.5	766	1 DPP4_HUMAN	P27487 h dipeptidy
34	987.5	24.5	766	2 Q53TN1_HUMAN	Q53TN1 homo sapien
35	984	24.4	765	1 DPP4_BOVIN	P81425 b dipeptidy
36	982.5	24.4	765	2 Q5R7G7_PONPY	Q5R7G7 pongo pygma
37	982	24.3	760	2 Q53TP5_HUMAN	Q53TP5 homo sapien
38	979	24.3	760	1 DPP4_MOUSE	P28843 m dipeptidy
39	977	24.2	760	1 SEPR_HOUSE	Q12884 homo sapien
40	975	24.2	761	2 Q8R492_RAT	Q8R492 rattus norv
41	974.5	24.2	755	2 Q91651_XENLA	Q91651 xenopus lae
42	972	24.1	761	1 SEPR_MOUSE	P97321 mus musculu
43	969	24.0	765	1 DPP4_FELCA	Q9N217 f dipeptidy
44	964.5	23.9	1077	2 Q4FP0_USTMA	Q4FP0 ustilago ma
45	962.5	23.9	878	2 Q75C14_ASHGO	Q75C14 ashbya goss

ALIGNMENTS

RESULT 1					
O42812 ASPOR					
ID	O42812_ASPOR	PRELIMINARY;	PRT;	771 AA.	
AC	O42812;				
DT	01-JUN-1998	(TREMBLrel. 06, Created)			
DT	01-JUN-1998	(TREMBLrel. 06, Last sequence update)			
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)			
DE	Prolyl dipeptidyl peptidase precursor (EC 3.4.14.5).				
GN	Name=dppIV;				
OS	Aspergillus oryzae.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;				
OC	Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.				
NCBI	TaxID=5062;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RA	Domas A., Van den Broek P., Affolter M., Monod M.;				
RT	"Characterisation of the prolyl dipeptidyl peptidase-encoding gene (dppIV) from the koji mould Aspergillus oryzae.";				
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AJ002369; CAA05343.1; -; Genomic_DNA.				
DR	HSSP; P27487; 1PFO.				
DR	MEROPS; S09.008; -.				
DR	GO; GO:0016020; C.membrane; IEA.				
DR	GO; GO:0004274; F.dipeptidyl-peptidase IV activity; IEA.				
DR	GO; GO:0016787; F.hydrolase activity; IEA.				
DR	GO; GO:0006508; P.proteolysis and peptidolysis; IEA.				
DR	InterPro; IPR001375; Peptidase_S9.				
DR	InterPro; IPR002469; Peptidase_S9B.				
DR	InterPro; IPR000379; Ser.estrs.				
DR	Pfam; PF00930; DPPIV_N; 1.				
DR	Pfam; PF00326; Peptidase_S9; 1.				
DR	PROSITE; PS00221; MIP; UNKNOWN_1.				
DR	Hydrolase; Signal.				
FT	SIGNAL	1	16	Potential.	
FT	CHAIN	17	771	prolyl dipeptidyl peptidase.	
SQ	SEQUENCE	771 AA;	87208 MW;	8AB0CB33797B3FC6 CRC64;	

Query Match					
Best Local Similarity					
Matches 733; Conservative					
Score 3886; DB 2; Length 771;					
Pred. No. 9.1e-240;					
Mismatches 19; Indels 4; Gaps 2;					
QY					
1	LDVPRKPHAPTGGSKRLTFNETVVKQAITPTSRVQWLSGAEDSGSYVAAEDGSLTIEN	60			
17	LDVPRKPHAPTGGSKRLTFNETVVKQAITPTSRVQWLSGAEDGS--LVRGGRQRSHH	74			
Db					
61	IVTNESRT--LIPADKIPGKEAFNFWIHPDLSSVLWASNHTKQYRHSFADYVQDVES	118			
75	REHRQRVTHAHPADKIPTGKEAFNFWIHPDLSSVLWASNHTKQYRHSFADYVQDVES	134			
QY					
119	LKSVPLMPDQEGDIQVAQNSPVGNTAFVRENDLYVWDNGTIVTRITDDGGPDMFHGVDPW	178			
135	LKSVPLMPDQEGDIQVAQNSPVGNTAFVRENDLYVWDNGTIVTRITDDGGPDMFHGVDPW	194			
Db					

```
QY 179 IYEEBILGDRVALWSPDGEYLAISFNETGVPTVTVQYVYMDNQEIAPAYPWELKIRYPK 238
DB 195 IYEEBILGDRVALWSPDGEYLAISFNETGVPTVTVQYVYMDNQEIAPAYPWELKIRYPK 254
QY 239 VSQTNPTVTVLSLNTASKEVKQAPIDAFESDGLIIGEVAMLTDTHTTTVAAKAFNRVQDOQ 298
DB 255 VSQTNPTVTVLSLNTASKEVKQAPIDAFESDGLIIGEVAMLTDTHTTTVAAKAFNRVQDOQ 314
QY 299 KVAVDTASNKATVISDRDGTGMDLNLMSKYGPIKPSDKADYIIDISDHSGWAHLYL 358
DB 315 KVAVDTASNKATVISDRDGTGMDLNLMSKYGPIKPSDKADYIIDISDHSGWAHLYL 374
QY 359 FVPSGGEPIPLTKGDWEVTSILSDQERQLVYVLSQHHSTERHLYSVSYSTFAVTPPLVD 418
DB 375 FVPSGGEPIPLTKGDWEVTSILSDQERQLVYVLSQHHSTERHLYSVSYSTFAVTPPLVD 434
QY 419 DTVAAYWASFSANSYYILTYGGDPVQYELTYTNSKPLRTITDNAKVLQIKDYALP 478
DB 435 DTVAAYWASFSANSYYILTYGGDPVQYELTYTNSKPLRTITDNAKVLQIKDYALP 494
QY 479 NITYPELPLPSGETLNVMLPSPGSPDKKYPILFTPYGGGAQOEVTXKQALNFKAYVA 538
DB 495 NITYPELPLPSGETLNVMLPSPGSPDKKYPILFTPYGGGAQOEVTXKQALNFKAYVA 554
QY 539 SDSELEYVTVTVNDRGTGKGRKFRSAVTRQGLLEAEDQIYAAQQAANIPIWDADHIGI 598
DB 555 SDSELEYVTVTVNDRGTGKGRKFRSAVTRQGLLEAEDQIYAAQQAANIPIWDADHIGI 614
QY 599 WGSFEGGYLTSKVLKDSGAFITVITAPVSDWRFYDSMYTERYMKLTSTNEEGYETSAV 658
DB 615 WGSFEGGYLTSKVLKDSGAFITVITAPVSDWRFYDSMYTERYMKLTSTNEEGYETSAV 674
QY 659 RKTDFGNKVEGFLQHGTDGDDNVHFQNSAALVDLLMGDGVSPKLSQWFTDSHGYSY 718
DB 675 RKTDFGNKVEGFLQHGTDGDDNVHFQNSAALVDLLMGDGVSPKLSQWFTDSHGYSY 734
QY 719 HGGGVFLYKQLARKLYQEKNRQTVLMHQWTKKDL EE 755
DB 735 HGGGVFLYKQLARKLYQEKNRQTVLMHQWTKKDL EE 771

RESULT 2
QSAZ42 EMENI PRELIMINARY; PRT; 773 AA.
ID Q5A242 EMENI PRELIMINARY;
AC Q5A242
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
GN ORFNames=AN6438.2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxId=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavsky L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepl Y., Collymore A., Cook A., Cooke P., Corum B., Dearellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardayon S., Gerre S., Graham L., Grand-pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamal M., Katat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt P., MacLean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrum J., Meneus L.,
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramaseamy U., Raymond C., Retta R., Rise C., Rogov P.,
```

```
RA Roman J., Schauer S., Schuback R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Testaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataran V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.
RT "Genome Sequence of Aspergillus nidulans.";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC !- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACD01000108; EAA58460.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 773 AA; 86954 MW; DEEB398491278E67 CRC64;

Query Match 74.9%; Score 3020; DB 2; Length 773;
Best Local Similarity 71.5%; Pred. No. 2.3e-184;
Matches 537; Conservative 107; Mismatches 101; Indels 6; Gaps 3;

QY 4 PRKPHAPTGESEKRLTFNETVVKQAIPTPSRSVOWLSGAEDGSVYAAEDGSLTIENIVT 63
DB 20 PWRPREPRAAGSKRLTFNETVISAALSPSSISVQVIATENDGDYVYQEDSGIKIESIVT 79
QY 64 NESRTLLPADKIPGKEAFNYIHPDLSSVLWASNHTKQYRHSFFADYVYQDVESLSKVP 123
DB 80 NRSOTIIVPAEKIPA-DAYSIIWISPDLSAVLWATNYTKQYRHSFFADYVYQDVETLEIVP 137
QY 124 LMPQDEGDIQYAOWSPGVNTIAFVRENDLYVMDNGTVTRITDDGGPDMFHGVPMWYBEE 183
DB 138 LVEDVMGDIQYAEWSPSGDSIAFVRGNLMTWSGTVTAITKCGGPDHGFVPMWYBEE 197
QY 184 ILGRYALWSPDGEYLAISFNETGVPTVTVQYVYMDNQEIAPAYPWELKIRYKVSQTN 243
DB 198 ILGRFALWSPDSELLAFLTFNETGVPTFTVQYFMDNQEIAPPYPRELDIRYPKVSQTN 257
QY 244 PTVTLISLNTASKVKQAPIDAFESDGLIIGEVAMLTDTHTTTVAAKAFNRVQDOQKVAV 303
DB 258 PTVKLNILQLSDNTVSTIPIIDVDFPSELIVGEVAMVTDHTTELAVKAFNRVQDSKVIV 317
QY 304 DTASNKATVISDRDGTGMDLNLMSKYGPIKPSDKD---AYYIDISDHSGWAHLYLFP 360
DB 318 ETASGETKIAHERDGTGMDLNLMSISYVGPLALGSGDASSAYVVDLSHSGWTHLYLFS 377
QY 361 VSGGEPIPLTKGDWEVTSILSDQERQLVYVLSQHHSTERHLYSVSYSTFAVTPPLVDDT 420
DB 378 TSGGDPILPTEGEWEVTSIVSDQERLVYVLSQHHSTERHLYSVSYRTFEITPLVDDT 437
QY 421 VAAWASFSANSYVILTYGGDPVQYELTYTNSKPLRTITDNAKVLQIKDYALPNI 480
DB 438 VEAYWSVFSKAGYIILTYAGSPVYQELYSVNTQAPLRLTNSAALIEKLEVALPNI 497
QY 481 TYFELPSPGETLNVMLPSPGSPDKKYPILFTPYGGGAQOEVTXKQALNFKAYVASD 540
DB 498 SYFELEIFSGEKLNVMLPSPGSPDKKYPVLPFTPYGGGAQOEVSXKQSLDFNAYIASD 557
QY 541 SELEYVTVTVNDRGTGKGRKFRSAVTRQGLLEAEDQIYAAQQAANIPIWDADHIGI 600
DB 558 PELEYVTVTVNDRGTGKGRKFRSAVTRQGLLEAEDQIYAAQQAANIPIWDADHIGI 617
QY 601 WSGGYLTSKVLKDSGAFITVITAPVSDWRFYDSMYTERYMKLTSTNEEGYETSAVRK 660
DB 618 WSGGYLTSKVLKDSGAFITVITAPVSDWRFYDSMYTERYMKLTSTNEEGYETSAVRK 677
QY 661 TDGFKNVVGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSQWFTDSHGYSYHG 720
DB 678 TDGFKNVVGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSQWFTDSHGYSYHG 737
QY 721 GGVLFLYKQLARKLYQEKNRQTVLMHQWTKK 751
DB 738 GNVFLYKQLARKLYQEKNRQTVLMHQWTKK 767

RESULT 3
```

Q4WPH9 ASPFU
ID Q4WPH9 ASPFU PRELIMINARY; PRT; 765 AA.
AC Q4WPH9;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Antigenic dipeptidyl-peptidase Dpp4.
GN ORFNames=Afu4g09320;
OS Aspergillus fumigatus Af293.
OC Eukaryota; Fungi; Ascomycotina; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillius.
NCBI_taxid=330879;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AF293;
RA Nierman W., Pain A., Anderson M.J., Wortman J., Kim H.Stanley,
RA Arroyo J., Bertram M., Abe K., Archer D.B., Bernaldo C., Bennett J.,
RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
RA Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,
RA Foster N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
RA Goldman G.H., Gomi K., Griffith-Jones S., Gilliam R., Haas B.,
RA Haas H., Harris D., Horluchi H., Huang J., Humphrey S., Jimenez J.,
RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,
RA Kumagai T., Latton A., Latge J.-P., Li W., Lord A., Lu C.,
RA Majors W.H., May G.S., Miller B.L., Mohamoud Y., Molina M., Monod M.,
RA Mouyna I., Mulligan S., Murphy L., O'Neill S., Paulsen I.,
RA Penalva M.A., Perlema M., Price C., Pritchard B.L., Quail M.A.,
RA Rabinowitch E., Rawlins N., Rajandream M.-A., Reichard U.,
RA Renaud H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA Takeuchi M., Tekata F., Turner G., Vasquez de Aldana C.R., Weidman J.,
RA White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
RA Machida M., Hall N., Barrell B., Denning D.W.,
RT "Genomic sequence of the pathogenic and allergenic filamentous fungus
Aspergillus fumigatus";
RL Submitted (MAY-2005) to the EMBL/GenBank/DDBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAHF01000005; EAL89855.1; -: Genomic DNA.
SQ SEQUENCE : 765 AA; 85863 MW; DA7BF150D80D686E CRC64;

Query Match 74.6%; Score 3010.5; DB 2; Length 765;
Best Local Similarity 72.2%; Pred. No. 9e-184;
Matches 544; Conservative 102; Mismatches 104; Indels 3; Gaps 2;

QY 1 LDVPRKPHAPTGBGSKRLTFNETVVKQAITPTSRVSQWLGAEDSGSYVAADGSLTTEN 60
DB 15 IDVPQPAPYPTGSGKKRLTFNETVVKRAISPAISVEIWTSEDGDYVVQDQGLKIQS 74

QY 61 IVTNSRTLIPADKIPTGKEAFNFWTHPDLSVLNASHTKOYRSPFDYVVDVESLK 120
DB 75 IVTNHTQTLPADKVP--EDAYSWIHPNLSSVLWATNYTKYRHSFYADYPIQDVQSMK 132

QY 121 SVPLMPDQEGDIOYAQWSPGVTGNTAFVRENDLVWDNGTVTRITDDGGDFMGHPDVIY 180
DB 133 LRPLAPDQSGDIQIAQWTPTGDAIALFRNNVFNWNTASTSQTINDGDPDLFNGVPDVIY 192

QY 181 EEELIGDRVALWFSPDGEYLAILSFNETGYPTTYOYTMDSNIAPAYPWELKIRYPKV 240
DB 193 EEELIGDRFALMFSPDGAYLAFLRFNETGYPTTYOYTMDSNIAPAYPRELELYPKV 252

QY 241 QTNPVTLLSLNTASKEVKQAIDPESTDLIIIGEVAWLTDTHTTVAAKAFNRVDQOKV 300
DB 253 QTNPTVELNLELRGTERTPVPIDAFDAKELIIGEVAWLTKGHVYAVKAFAFRNVDRQKV 312

QY 301 VAVDTASKATVISDRDGTGMDLNLSMKYIGPIKPSDKDAYIIDSDSHGWAHLXLP 360
DB 313 VAVDVASLRKTSERDGTGMDLNLSMAIYGPIGES-KEEYIIDSDSHGWAHLXLP 371

QY 361 VSGGERIPLTKGWDEVTSILSIDQERQLVYLLSTOHSTERHLYSVSYSTFAVTEPLVD 420

Db 121 SVPLMPODEGDIQVAQMSPVGNTAFVRENDLYVWDNGTVTRIIDDGPDPMFHGVPDWIY 180
Qy 133 LRPLAPOSGDIIQVAQMSPTGDAIAFVGNVNVFWMTNASTSQIINDGPDPLFNGVDPWIY 192
Db 181 EEEILGRYALWFSPDGEYLAYLSFNETGVPTTYVQYMDNQEIAPAYPMELKIRYPKVS 240
Qy 193 EEEILGRFALWFSPDGAYLAFLAFNFTGVPTTVPYMDNEEIAAPPYPRLELRYPKVS 252
Db 241 QTNPTVTLNLINIAKEVKQAPIDAPESDILIEGEVWLTDTHTTVAKAFNRVODQKV 300
Qy 253 QTNPTVELNLELRGTPTPIDAFDAKELIIEGEVWLTGKHVDVAVKAFNRVODQKV 312
Db 301 VAVTASNAKATVSDRGTGDLNLLSMKYGIPKPSDKDAYYDIDSHGSAWHLVLP 360
Qy 313 VAVDVASLRSKTIISERDGTGDLNLLSMAYIGPAGES-KEEYIYDIDSQSAWHLVLP 371
Db 361 VSGGEPIPLTKGDMVEVTSILSIDORQLVYYLSTQHHSTERHLYSVSYSTFAVTPLVDDT 420
Qy 372 VAGGEPIALTKEWEVNLISIDKPRQLVYFLSTKHHSTERHLYSVSMKTEITPLVDDT 431
Db 421 VAAVMSAFSANSQYIITYGGPDVYQELIYTTNSTKPLRTITDNKVLKQIKOYALPNI 480
Qy 432 VPAVMSAFSQQGGYIILSYRGPDVYQDLVAINSTAPLRTITNSAALNALKEYTLFNI 491
Db 481 TYFELPLSGEGLNVMQRLPGEFSDKKYPILFPTYPGGPGAQEVTKRQALNFKAYVASD 540
Qy 492 TYFELALPSGETLNMVQRLPVKFSKKYPVLFPTYPGGPGAQEVSKPQALDFKAYIASD 551
Db 541 SELEVVTWVNRGTGFGKGRFSRATVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGIWG 600
Qy 552 PELEVITWVNRGTGYKGRFRCQVASRLGELEAADOVFAAQQAALPYVDAQHIAIWG 611
Db 601 WSGGYLTSKVLKDSGAFPTLGIVITAPVSDWRVFDVSMYTERYMTKLTLSNBSGYETSAVRK 660
Qy 612 WSGGYLTGKVIETDSGAFSLGVQAPVSDWRVFDVSMYTERYMTKLTLSNAAGYNASAIRK 671
Db 661 TDGPKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKELHSQWFTSDHIGISVHG 720
Qy 672 VAGYKXNRGGLVLIQHTGDDNVHFQNSAALVDLLMGDGVTPKLVQVQWFTSDHIGIRYHG 731
Db 721 GGVFLYKOLARKLYOEKNRQTOQLMHQWTKKDL 753
Qy 732 GNVFLYRQLSKLEIZEKKRKEGEAHQSKSV 764

RESULT 5

Q5J6J3_TRIRU PRELIMINARY; PRT; 775 AA.
ID Q5J6J3_TRIRU PRELIMINARY; PRT; 775 AA.
AC Q5J6J3_10-MAY-2005 (Tremblrel. 30, Created)
DT 10-MAY-2005 (Tremblrel. 30, Last sequence update)
DT 10-MAY-2005 (Tremblrel. 30, Last annotation update)
DE Dipeptidyl-peptidase IV.
OS Trichophyton rubrum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Onygenales; Arthrodermataceae; mitosporic Arthrodermataceae;
OC Trichophyton.
OX NCBI_TaxID=5551;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15632434; DOI=10.1099/mic.0.27484-0;
RA Monod M., Lechenne B., Jousson O., Grand D., Zaugg C., Stocklin R.,
RA Grouzmann E.;
RT "Aminopeptidases and dipeptidyl-peptidases secreted by the
RT dermatophyte Trichophyton rubrum.";
RL Microbiology 151:145-155 (2005).
DR EMBL; AY497021; AAS7665.1; -; Genomic_DNA.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004274; F:dipeptidyl-peptidase IV activity; IEA.

DR GO; GO:0004185; F:serine carboxypeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002471; Pept_S9_AS.
DR InterPro; IPR001563; Peptidase_S10.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002469; Peptidase_S9B.
DR InterPro; IPR000379; Ser_estrs.
DR Pfam; PF00930; DPPIV_N; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
DR PROSITE; PS00708; PRO_ENDOPEP_SER; UNKNOWN_1.
KW Hydrolase; Transmembrane.
SQ SEQUENCE 775 AA; 88021 MW; 157774F128C6BFDB CRC64;

Query Match 64.7%; Score 2610.5; DB 2; Length 775;

Best Local Similarity 61.6%; Pred. No. 3.5e-158;

Matches 463; Conservative 134; Mismatches 148; Indels 7; Gaps 4;

Qy 4 PRKPAPTGGSKELTFNETVVKQAITPSTRSVOVLSGAEDGSVYAAEDGSLTIENIVT 63
Db 19 PREPRSPGTGGNKLLTYKECVPRATISPRSTSLAWINSEEDGRYISQSDGALLQNIIVT 78
Qy 64 NESRTLPADKIPGKEAFNYWIHPDLSSVLWASNHTKQYRHSFPFADYVYQDVESLSKVP 123
Db 79 NTNKLVAADKVPKG--YYDYWFKPDLSAVLWATNYTKQYRHSYFANYFILDIKKGSILTP 136
Qy 124 LMPDQEGDIQVAQMSPVGNTAFVRENDLYVWDNGTVTRIIDDGPDPMFHGVPDWIYEE 183
Db 137 LAQDAQGDIQVAQMSPMWNSIAYVRXNDLYIWNNGTKRITENGSGPDIFNGVDPWYBEE 196
Qy 184 ILGRYALWFSPDGEYLAYLSFNETGVPTTYVQYMDNQEIAPAYPMELKIRYPKVSQTN 243
Db 197 IFGRFALWFSPDGEYLAYLRFNETGVPTTYPIYKXKQKIAPAYPRELEIRYPKVSAN 256
Qy 244 PTVTLSLNLINIAKEVKQAPIDAPESDILIEGEVWLTDTHTTVAKAFNRVQDQKVAV 303
Db 257 PTVOFHLNLINIASSQETIPVTAFFENDLVIEGEVWLSGHDVAVYAFNRVQDREKIVSV 316
Qy 304 DTASNKATVISDRDGTGDLNLLSMKYIGIPKPSDKDAYYDIDSHGSAWHLVLPVSG 363
Db 317 KVESKESKVIIRERDGTGDLNLLSMSYIGNVGKE--YYVDISDASGWAHIYLPVVG 373
Qy 364 GEPTPLTKGDMVEVTSILSIDORQLVYYLSTQHHSTERHLYSVSYSTFAVTPLVDDTVAA 423
Db 374 GKEIALTKEWEVVAAILKVDTKKLIYFTSKYHSTHRYVSVSYDTKVMTPPLVNDKEAA 433
Qy 424 YWSAFSANSQYIITYGGPDVYQELIYTT-NTSKPLRTITDNKVLKQIKOYALPNIY 482
Db 434 YITASFSKAGYYIILSYQGNVPYQELYSTKDKKPLKTIISNDALLEKLEKEYKLKPVSF 493
Qy 483 FELPLSGEGLNVMQRLPGEFSDKKYPILFPTYPGGPGAQEVTKRQALNFKAYVASDSE 542
Db 494 FEIKLPSGETLNVKQRLPPNPNPHKKYFVLFTYPGGPGAQEVSAWNSLDKFSYITSDPE 553
Qy 543 LEYVTVTVNDRGTGFGKGRFSRATVTRQLGLLEAEDQIYAAQQAANIPIWDADHIGIWS 602
Db 554 LEYVTVTVNDRGTGYKGRKFSRASAVALGFLGLEAQQVFAAKEVLKXNRWADKDHIGIWS 613
Qy 603 FGGVLTSTKVLKDSGAFPTLGIVITAPVSDWRVFDVSMYTERYMTKLTLSNBSGYETSAVRKT 662
Db 614 YGGFLTAKTLETDGSGVFTFGISTAPVSDFRLYDYSMYTERYMTKLTLSNBSGYETSAVRKT 673
Qy 663 GPKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKELHSQWFTSDHIGISYHGG 722
Db 674 GFKNLKGHYLLQHGTDGDDNVHFQNAALVSLNLLMGVGTADKLTTQWFTSDSHGIRYDMS 733
Qy 723 VFLYKQLARKLYOEKNRQTO-Q-VLMHQWTKDL 753
Db 734 TVQYKQLSKMYYDQKRRPESPPMHQWSKRVL 765

RESULT 6

Q51ZP7_MAGG

Q512P7_MAGGR PRELIMINARY;		PRT; 778 AA.		Matches 443; Conservative 124; Mismatches 181; Indels 5; Gaps 4			
ID	Q512P7						
AC	13-SEP-2005 (TrEMBLrel. 31, Created)						
DT	13-SEP-2005 (TrEMBLrel. 31, Last sequence update)						
DT	13-SEP-2005 (TrEMBLrel. 31, Last annotation update)						
DE	Hypothetical protein.						
GN	ORFNames=MG05989.4;						
OS	Magnaporthe grisea 70-15.						
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;						
OC	Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.						
OX	NCBI_TaxID=242507;						
RN	[1]						
RP	NUCLEOTIDE SEQUENCE.						
RC	STRAIN=70-15;						
RA	Birren B., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,						
RA	Ait-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,						
RA	Arachchi H., Armbruster J., Bachantsang P., Baldwin J., Barry A.,						
RA	Bayul T., Blitshteyn B., Bloom T., Blye J., Boguslavskiy L.,						
RA	Borowsky M., Boukhgaiter B., Brunache A., Butler J., Calixte N.,						
RA	Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,						
RA	Collimore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,						
RA	David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,						
RA	Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,						
RA	Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,						
RA	Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,						
RA	Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,						
RA	Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,						
RA	Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,						
RA	Jaffe D., Jones C., Kamal M., Kamat A., Kamyselis M., Karlsson E.,						
RA	Kellis C., Kieu A., Kinsner P., Kodira C., Kulbokas E., Labutti K.,						
RA	Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,						
RA	Lindblad-toh K., Liu X., Lokitsang T., Lokitsang Y., Lucien O.,						
RA	Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,						
RA	Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,						
RA	McCarthy M., McDonough S., McGhee T., Meldrim J., Meneus L.,						
RA	Mesirov J., Mihalev A., Mihova T., Mikkelson T., Mlenga V., Moru K.,						
RA	Mozes J., Mulrain L., Munson G., Naylor J., News C., Nguyen C.,						
RA	Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,						
RA	Norbu N., O'donnell P., Okawo O., O'leary S., Omotosho B.,						
RA	O'Neill K., Omsan S., Parker S., Perrin D., Phunkhang P., Pigani B.,						
RA	Purcell S., Rachupka T., Ramaamy U., Rameau R., Ray V., Raymond C.,						
RA	Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,						
RA	Rutman M., Schupbach R., Seaman C., Settippalli S., Sharpe T.,						
RA	Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,						
RA	Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,						
RA	Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,						
RA	Tenzing P., Tesfaye S., Theodore J., Thoultsang Y., Topham K.,						
RA	Towey S., Tsamila T., Tsomo N., Vallee D., Vassiliev H.,						
RA	Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,						
RA	Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,						
RA	Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,						
RA	Zimmer A., Zody M., Lander E.;						
RT	"The genome sequence of Magnaporthe grisea.";						
RL	Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.						
RN	[2]						
RP	NUCLEOTIDE SEQUENCE.						
RC	STRAIN=70-15;						
RA	Dean R., Mitchell T., Brown D., Pan H., Thon M.;						
RL	Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.						
RN	[3]						
RP	NUCLEOTIDE SEQUENCE.						
RC	STRAIN=70-15;						
RA	Zhu H., Blackmon B.;						
RL	Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.						
CC	-1- CAUTION: The sequence shown here is derived from an						
CC	EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is						
CC	preliminary data.						
DR	EMBL; AACU01000849; EAA52861.1; -; Genomic_DNA.						
KW	Hypothetical protein.						
SQ	SEQUENCE 778 AA; 86791 MW; FA19726C109BA278 CRC64;						
Query Match 59.8%; Score 2411.5; DB 2; Length 778;							
Best Local Similarity 58.8%; Pred. No. 1.9e-145;							

RESULT 7

Q4HM6 GIBZE

ID Q4HM6 GIBZE PRELIMINARY; PRT; 779 AA.

AC Q4HM6;

DT 13-SEP-2005 (TrEMBLrel. 31, Created)

DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)

DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)

DE Hypothetical protein.

GN ORFNames=FGI0982.1;

OS Gibberella zeae PH-1.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

OX NCBI_TaxID=229533;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=PH-1;

RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,

RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,


```
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Miengra V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nilsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Testfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RA "Fusarium graminearum genome sequence.";
RA Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACM0100457; EAA75308.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 779 AA; 87757 MW; 3EC8D468C633F192 CRC64;

Query Match 59.5%; Score 2399.5; DB 2; Length 779;
Best Local Similarity 57.0%; Pred. No. 1.1e-144;
Matches 433; Conservative 140; Mismatches 179; Indels 7; Gaps 5;

QY 1 LDVPRKHAPTEGSGKRLTFNETVVKQAITSRSVQVLSGAEQSGVYAAEDGSLTIEN 60
DB 20 VDPKRKHQPTGEGDRLTLTFNETSPSKALRPSSISVDHSSAGGDNVILSDGDLVLED 79
QY 61 IVTNESRLTIPADKIPTGKEAFNYWIHPDLSSVLWASNHTKOYRHSFPADYVVDVSLK 120
DB 80 ILSEKTKFTVADKL P--KDLHEYISGDAKKVLVASNTKOYRYSFADYFVLVDSEGE 137
QY 121 SVPLMPDQEGDIOVAQWSPGVNNTAFVRENDLYV-WDNGTTRITDDGPPMFHGVDPWI 179
DB 138 SEPLVDNQVGIQVAQFAPNGDAVAFVRDNDLFIIRSDGKVDQITSDGPPMFHGVDPWV 197
QY 190 YEEETLGRYALWSPDGEYLAYLSFNETGPTVTVQVYMDNOETAPAYPWEKIRYPKV 239
DB 198 YEEETLGRYALWSPDGEYLAYLSFNETGPTVTVQVYMDNOETAPAYPWEKIRYPKV 257
QY 240 SQTNPVTLSLINTASKEVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDOOK 299
DB 258 GSKNPTVELNLDVSEGEYKVPVVEAFPEELIIGEVAVWTKDHSALIFRAFNRVQDOQA 317
QY 300 VVAVDTASKATVISDRDGTGWLNDLLSMKYIGIKPSDDKDAYIDISDHSWAHLVLF 359
DB 318 HVVVNPETLKSNTVREDRDGDGWLSEHTLSISFVGPLSGSGNSYVVDNDQDGNHLYL 377
QY 360 PVSGCEPIPLTKGDMEVTSILSIDQEROLVYVLSQTHSTERHLYSVSYSTFAVTPLVDD 419
DB 378 PVNGGEALQLTSGEVEVSTILNIDTAELVLPVQAKRHSSTERHLYSVSYSTFAVTPLVDD 437
QY 420 TVAAWYWSASFANSYGIYILTYGGDPVPYQELTYTNST-KPLRTITDNTAKVLEIQIKDALP 478
DB 438 SVPGYWSASFSSGGYIYLHYQGPDPVPYQELTYTNSTSKPVRTLVNDKQFYKNITQYNLP 497
QY 479 NITYPELPLPSGETLNVQRLPPGSPDKKYPILFTPGCGAQQEYVTRKQALNFKAYVA 538
DB 498 NITYPELHPDGYTLNVQQLPPNFNASHKYVPLFTPGGPNRSQVAKSFQALGWKQYIS 557
QY 539 SDELEYVYVTVNDNGTGFPGKFKFSVATROLGLEAEADQIYAAQQAANI-PWIDADHIG 597
DB 558 SDFELQYVVYTVNDNGTAMRGKRYKRSWTSQGLKLEPLDQIWAASEQLTSMFDPVNAEKIG 617
```

```
QY 598 IWCWSGGYLTSKVLEKDSGAFTLVITAPVSDWRFYDMSMYTRYMYKTLSTNBEGETSA 657
DB 618 MWGWSVGGYLTAKTIADSGVFSGLITAPVSDWRFYDMSMYTRYMYKTLSTNBEGETSA 677
QY 658 VRKTDFGNVEGFLIQHGTGDDNVHFQNSAALVLLMGDGVSPKLSHQWFTDSDHGIS 717
DB 678 VHDVSGFKNIAGQFSVLHGTGDDNVHQAAMIDLLVSGSVSPKMKMFAFTDSDHSIV 737
QY 718 YHGGGVFLYKQLARKLYQEKNRQ--TQVLMHQWTKDLE 754
DB 738 YNGASVMIYKYLTLARLYDEVKQPKAKALHTHONKRIE 776

RESULT 8
Q4IAY7 GIBZE
ID Q4IAY7 GIBZE PRELIMINARY; PRT; 911 AA.
AC Q4IAY7;
DC 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=FG05621.1;
OS Gibberella zeae PH-1.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocorynetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=229533;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PH-1;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Miengra V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nilsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Testfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RA "Fusarium graminearum genome sequence.";
RA Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACM01000231; EAA75192.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 911 AA; 102542 MW; B30DFE00EE04AB59 CRC64;

Query Match 34.1%; Score 1376; DB 2; Length 911;
Best Local Similarity 39.9%; Pred. No. 4.1e-79;
Matches 317; Conservative 125; Mismatches 285; Indels 68; Gaps 25;

QY 2 DVPRKHAPTEGSGKRLTFNETVVKQAITSRSVQVLSG--AEDGSY--VYAAEDGSLT 57
DB 117 DTEHDPDAATRGSGKTVTMDQ-LFDGSMQAKSHSISWIAGPDGEDGLLEVGAASEKPYIV 175
QY 58 IENIVTNES-----RTLIPADKIPTGKEAFNYWIH-----PDLSSVLWASNH 99
DB 176 VEDIRKDKENRAPINTEIKASKSRITMKD-NFFVHDGNOYTFDMNRPDPLKKVLGVNKK 234
QY 100 TKQYRHSFFADYVVDVQDVESLKSVPMLPQ-EGDIOYAQWSPGVNNTAFVRENDLYVW--- 155
```

Db 235 KKNWHSFSAIFILFVETQAEPLVPGKPDARIQLATWSPKSDAVSVTQDNLLYIRRLD 294
QY 156 DNGTTRITDDGPDPMFCHGVPWYEEELGRDYLALWSPDGEYLALYLSFNETGTYTIV 215
Db 295 DKQNVQITKDGPGFYFGIPDMWYEEVFGSRATWMSDEGKYLFLRTNETGVPFPV 354
QY 216 QYVM-----DNOETAPAPWELKIRYKPVQSTNPTVTLISLNIASKEVKQAPID-APES 268
Db 355 QPFIKPSGTDEKEEAYPEVQIKYKPGAGHNPVVDLLFYDIKQVFSVDIDGAFAD 414
QY 269 TDLIIGEVAWLDTHTTVAAKAFNRVQDQKVAVDATASNKATVIS--DRDGTD-GWLDN 325
Db 415 DRIINLLWAGD--KAVVETNRVSDILKVVLIDVTSRKGKTNTVTDVXIDGQWPEI 471
QY 326 LLSMKYI--GPIKPSDKDAYIDIDSHSGWAHL-YLPVSGGEPILPTKGDMEVTSILS- 381
Db 472 SHTYAIIPADPANGRDHOG-YVDSYLHEGYHLAYFTPLDNPPIMLTGGEWEIDDAPSA 530
QY 382 IDQERQLVYLTSTOHHSTERHLYSVSYSTFAVTPLVDDTVAAYWASFSANSYVILTYG 441
Db 531 IDLANNLVTFIAKSSIOGHVSVKLDGTEKLEALTDPEYAYDASFSKSGFVLLSYR 590
QY 442 GPDVPYQELTYTNTST--KPLRTITDNNAKVLQIKOYALPNITYFELPLPSGETLANMORL 499
Db 591 GPKVPSQKVIPTSSVHDYEHVIEDNAELAEARSHELPVMKYGTLDLGDGIWVNVERR 650
QY 500 PPGFSPDKKYPILPTYPGFGAQEVTKRQALNFKAYVADSELEYVTVTVNDRGTGPKG 559
Db 651 PPHFDKKKYPVLFQOYSGFGSQSVTKRF-AVDFAQYVA--SALGYLVVTVDPRTGFLG 707
QY 560 RKFRSAVTQLGLAEADQIYAAQQAANIPTWIDADHIGWMSFGGYLTSKVLKDSG-A 618
Db 708 RHRVVRVRSQGLVESQDHIHAAQSPASRPYDVAERLAWGSHYGGFTTLKLEQDAGRT 767
QY 619 FTLGVTAPVSDWRFYDSMYTERYMKTLSTNEEGYETSVAVRTKDGPKNVEGG---FLIOH 675
Db 768 FSYGMAVAVPTDWRFYDSIYTERYMTPOENPGYDMSMIANATAL---GGNKRFLVMH 823
QY 676 GTGDDNVHFQNSAALVDLLMGDGVSPKELHSHWFTDSHGISHYGGGVFLYKQARKLYQ 735
Db 824 GVADDNVHFQNSLTLLDLSLDLAGVENYDVHV--FPDSDHSIYFHNNGRIYVDKLRNWLIN 881
QY 736 EXNRQTOVLHQWTK 750
Db 882 AFN-----GEWLK 889

RESULT 9

Q96VT7 ASPNG PRELIMINARY; PRT; 901 AA.
AC Q96VT7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Dipeptidyl aminopeptidase type IV (EC 3.4.14.5).
GN Name=dap;
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=CBS 120.49;
RA Jalving R., ter Veen W.J., Schaap P.J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ278532; CAC41019.1; -; Genomic_DNA.
DR HSSP; P27487; 1pFQ.
DR GO; GO:0016020; C-membrane; IEA.
DR GO; GO:0004177; F-aminopeptidase activity; IEA.
DR GO; GO:0004274; F-dipeptidyl-peptidase IV activity; IEA.
DR GO; GO:0016787; F-hydrolase activity; IEA.
DR GO; GO:0006508; P-proteolysis and peptidolysis; IEA.

DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002469; Peptidase_S9b.
DR InterPro; IPR003379; Ser_eatrs.
DR Pfam; PF00930; DPEIV_N; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Aminopeptidase; Hydrolase.
SQ SEQUENCE 901 AA; 101256 MW; 3A519512C903D784 CRC64;
Query Match 32.1%; Score 1295; DB 2; Length 901;
Best Local Similarity 38.4%; Pred. No. 6.2e-74;
Matches 294; Conservative 133; Mismatches 287; Indels 52; Gaps 26;
QY 11 TGEGSKRUTFNETHVVKQAITPSTRSVQWLSG--AEDGSVYAAED---GSLTIENI---- 61
Db 122 SSSSGKPVTLQVLTGQWL-PRSHAVSWIAGNGEDGLLVEQEGQKGKGLVDRSRK 180
QY 62 ---VTNERTLIPADKIPT-KGEAF--NYVTHPDLSSVLWASNHTKQYRHSFADYVVD 115
Db 181 GDATSQESRVLMEKAIVQVDGRTIPFVSTWSPFNKVKLLLSERKNWRHSFTGKWIIFD 240
QY 116 VESLKSVPIMP--DQSGDIQYQWSPVGNITAFVRENDLYV--WDNGTTRITDDGPDMP 172
Db 241 VATQTAQPLDFSNPGRVQLAIWSPSDMVAFVRDNNLYLRLLSSKEVVPITKGGADLF 300
QY 173 HGVPMWYEEELGRDYLALWSPDGEYLALYLSFNETGTYTIVQYVM---DNOETAPA-- 227
Db 301 YGIPDMWYEEVFGSNVSTWMSGDGKYVAFRLTNETAVPEFFVQYLYLSRPSGKRPPLGLE 360
QY 228 YPWELKTRYKPVQSTNPTVTLISLNIASKEV--KOAPIDAFESTDLIIGEVAWLDTHT 284
Db 361 DYPEVREIKYKAGAPNPVMSLQFYDVEKQEVFSIEAP--DFEDDDRIVIEIWW--GTG 417
QY 285 TVAAKAFNRVQDQKVAVDATASNKATV-----ISDRDGTGMDLNLMSKMYGIPKPSD 339
Db 418 KILVPAWRESVLAFLVFDTKARTSKLVRTENVADIDG--GWERTQYTWFI-PADPSN 474
QY 340 KDAY--YIDISHSWAHL-YLPVSGGEPILPTKGDMEVTSI-LSIDERQLVYVLSIQ 395
Db 475 GRPHDGYLDTVTHEGYEHLGYFTPLDNGSEPIILLTQGEWEVVDAPTAVDLRKGVYFISTK 534
QY 396 HHSITERHLYSVSYSTFAVTPLVDDTVAAYWASFSANSYVILTYGGPDVPYQELVT--T 453
Db 535 ESPTERHLYQVNLDSGNLKLPTDTSKPGYDYVSFSGHTGYALLSYKGPSIPQAIWNTET 594
QY 454 NSTKPLRTITDNNAKVLQIKOYALPNITYFELPLPSGETLANMQRLLPFGSPDKKYPIIF 513
Db 595 DELKTEETIEDNAGLARWVDSYALPTEIYQNTVI-DGFTLVVERPPHFNPAKYPVLP 653
QY 514 TPYGGFGAQEVTKRQALNFKAYVADSELEYVTVTVNDRGTGFKGRKFRSAVTROGLLL 573
Db 654 YLYNGPRSQTVDRKF-SIDFQSYVA--SSLGVIYVTVDGRGTGFSGRKTRCIVRGNLGY 710
QY 574 EADQIYAAQQAANIPTWIDADHIGWMSFGGYLTSKVLKDSG-AFTLGVTAPVSDMR 632
Db 711 EAYDQITTANLWGERKPYVDTERMSWGSYGGFMTLKLEQDAGQTFQYGMVAVPTDNR 770
QY 633 FYDSMYTERYMKTLSTNEEGYETSVAVRTKDGKNGVEGFLIOHGTGDDNVHFQNSAALVD 692
Db 771 HYDSIYTERYMTTPAHNPNGYNTSITDMTALQQTV-RFLVINGASDDNVHIGNTLVLVD 829
QY 693 LLMGDGVSPKELHSHWFTDSHGISHYGGGVFLYKQARKLYQAEKN 738
Db 830 KLDLAGVQNYDLH--FYPDSDHSINFNHNAHRMVMYERLSSWLNVAFN 873
RESULT 10
Q7SI80 EMENI PRELIMINARY; PRT; 874 AA.
AC Q7SI80;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Dipeptidyl aminopeptidase.

GN Name=etel3;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;
RX MEDLINE=22830407; PubMed=12949156; DOI=10.1099/mic.0.0.C0119-0;
RA Dyer P.S., Paolletti M., Archer D.B.;
RT "Genomics reveals sexual secrets of Aspergillus";
RL Microbiology 149:2301-2303(2003).
CC -!- MISCELLANEOUS: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ third party annotation (TPA) entry.
DR EMBL; BK001296; DAA01787.1; -; Genomic_DNA.
DR HSSP; P27487; 1NU6.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004177; F:aminopeptidase activity; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004274; F:dipeptidyl-peptidase IV activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001375; Peptidase S9.
DR InterPro; IPR002469; Peptidase S9.
DR InterPro; IPR000379; Serine.
DR Pfam; PF00930; DPIP_N; 1.
DR Pfam; PF00326; Peptidase_S9; 1.
KW Aminopeptidase.
SQ SEQUENCE 874 AA; 98282 MW; 737152B6D76B6C80 CRC64;

Query Match 32.0%; Score 1289; DB 2; Length 874;
Best Local Similarity 38.2%; Pred. No. 1.4e-73;
Matches 294; Conservative 123; Mismatches 280; Indels 72; Gaps 24;

QY 11 TGEGRKLTFTNETVVKQAITPTSRVQWLSG--AEDGSVYVAED---GSLTIENI---- 61
DB 132 TSSG-KPITLQ-Q-VLSGVFLPRGHAISSVAGDGEDGLIERGEDDEAGYLRIINDIRQDG 189
QY 62 -----VTNESRTLIPADKIPTGKEAFNYWIHPDLSSVLWASNHTKQYRHSFFAD 110
DB 190 KVNRLVMQKPTVGVDRITKPSATRPS-----PDLKKVLIISNQEKNRWHSFTAS 239
QY 111 YVQDVESLKVPLMPDQ-EGDIOVAQSWPVGNTIAFVRENDLYV--WDNGTVTRITDDG 167
DB 240 YWIFDVETQTAPELPLDNNIDGRVQLWSPKSDAIAFVRDNNLYRLKLSSEKRVVPIITKDG 299
QY 168 GPDMPHGVDPMTYEEIILGDRVALWFSPDGEVLAVLSFNETGVPTTYQYYMDN----- 221
DB 300 GEQLFYGVDPDWYEEVEFSGNSVTWSEDSQIAFIRTNESAVPEFPVQYFLSRPSGKKP 359
QY 222 QEIAPAYPELKIYRYPKVSQTNPTVTLSSLNIAKSKVQAPI-DAFESTDLIIIGEVAMLT 280
DB 360 QPGLNYPEVREIKYKAGAPNFFVNLQFYDVEQGEVFSVDTDPDDFDDDRLLIIEVIA- 418
QY 281 DTHITVAAKAFNRVQDQKVAVDVTSKATV-----ISDRDGTGDLNLLSMKYIGPI 335
DB 419 -AKGKVLVTRTNRSDILKVLVDVTSRESKLIHQIDISELQD--GWPEQTSVRFI--- 472
QY 336 KPSDKDA-----YVIDSDHSGWAHL-YLFPVSGGEPIPLTKGDWEVTSI-LSIDQERQL 388
DB 473 -PADPDKGRPFPGYLDVTVHGGDHLAVFTPLDNEPIMLTSGEVEVDVADTAVDLTRGL 531
QY 389 VYVLTSTQHHSTERHLYSYSTFAVTPVLVDDTVAAAYWSASFANSYGLTYTGGDPVYQ 448
DB 532 VYFIATKEAPTERHLYRVLDSGLTDLTDTSPQGVYSVFSFGDAGYALLSYQGSIPWQ 591
QY 449 ELYTT--NSTKPLRITDNKVLQEIQYALNITYFELPLPSGSLTNMQRLPFGSPD 506
DB 592 SIISTEGEKTTTLRIEDNTDLSKLVQAVALPTENYQNIITI-DGFTLQVVERRPHEFNP 650
QY 507 KVPILFTPEYGGFGAQEVTKRQALNFKAYVSDSELEYVTWVDNRGTFGKGRFGRSAV 566
DB 651 RKYPVLFLHYGGFGSGQTVDRRFN-VDFQSYVA--ASLGIYVTVVDGRGTGFGIGRAACII 707

QY 567 TROLGLLEAEADQIYAAQQAANIPIWDADHIGIWGSGFGYLTSKVLKDSG-AFTLGLVIT 625
DB 708 RGMIGHYEADQIATATKNWAQKPYDSESRVAIWGSGYGGFWTLKLEQDAGEITFOYGMV 767
QY 626 APVSDMRFYDSMYTERYMKTLSTNEEGYETSVAVRKTDGFKNVGEGFYLQIHGTGDDNVHFQ 685
DB 768 APVDMRFYDSVYTERYMHPTQHNPTGYDNTSISDMAALHN-NVRFVLVIHGASDDNVHIQ 826
QY 686 NSAAVLDLLMGDGVSPKLSQMFPTSDHGHSIHGGVGVFLYKQLARKLY 734
DB 827 NTLTLIDKL--DLASVQNVYDVHFYDSDHSIFFHNAHTMVYERMSPIFY 873

RESULT 11

Q5B934 EMENI PRELIMINARY; PRT; 880 AA.
ID Q5B934; Q5B934; Q5B934;
AC Q5B934;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
GN ORFNames=AN2946.2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArrellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Fato S., Ferreira P., Fitzgerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaife D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Mienga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schupbach R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Testaye S., Theodore J., Topham K., Travers M.,
RA Vasiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Genome Sequence of Aspergillus nidulans";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACD0100051; EAA63517.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 880 AA; 99015 MW; BF0FFB0CDOEA81BC CRC64;

Query Match 31.8%; Score 1284; DB 2; Length 880;
Best Local Similarity 38.4%; Pred. No. 3e-73;
Matches 293; Conservative 121; Mismatches 277; Indels 72; Gaps 24;

QY 11 TGEGRKLTFTNETVVKQAITPTSRVQWLSG--AEDGSVYVAED---GSLTIENI---- 61
DB 132 TSSG-KPITLQ-Q-VLSGVFLPRGHAISSVAGDGEDGLIERGEDDEAGYLRIINDIRQDG 189
QY 62 -----VTNESRTLIPADKIPTGKEAFNYWIHPDLSSVLWASNHTKQYRHSFFAD 110
DB 190 KVNRLVMQKPTVGVDRITKPSATRPS-----PDLKKVLIISNQEKNRWHSFTAS 239
QY 111 YVQDVESLKVPLMPDQ-EGDIOVAQSWPVGNTIAFVRENDLYV--WDNGTVTRITDDG 167

```
Db 240 YWIPDVETQASPLDNNIDGRVQLAWSPKSDAIAFVRDNNLYLKLSSERWVPIITQD 299
QY 168 GPMDFGVDPDWIYEELGDRVALWFSPPGEVLAFLSNFTGVPYTYVQYMDN----- 221
Db 300 GBLQFYGVDPDWIYEELVSGNSVTWMSDGSQAFIRTNESAVPEFFVQYFLSRSGKKP 359
QY 222 QBIAPAYPWLKIRYPKVSQTNPVTLSLNTASKEVQAPI-DAFESTDLIIIGEVAWLT 280
Db 360 QGLENYPVRIKPKKAGAPFVNLQFYDVEQGEVFSVDPDPDFDDDLRIIEVIA- 418
QY 281 DTHTTVAAKAFNRVQDQKVAVDATSNKATV-----ISDRDGTGWLNDLLSMKYIGPI 335
Db 419 -AGKVLVTRTNKRESILKVLVDVTESRESKLIRIQDISLQD--GWVEPTQSVRFI--- 472
QY 336 KPSDKDA-----YYIDISHSQWAHL-YLFPVSGGEPILTKGDWEVTSI-LSISOEROL 388
Db 473 -PADPKGRPPFGYLDVTVVHEGYDHLAYFTPLDNPPEPIMLTSGEWEVDAPTAVDLTRGL 531
QY 389 VYVLSQHHSTERHLYSVSYSTFAVTPLVDDTVAAYWSASFANSGYIILTYGGDPVPYQ 448
Db 532 VFIAIKEAPTEHLYRVLDGSDLPLTDTSQPGYYSVSPDGCAGYALLSYQGSIPWQ 591
QY 449 ELYTT--NSTKPLRTITDQNAKVLQIKDYALPNITYFELPLPSGETLNVQRLPPGSPD 506
Db 592 SIISTEGEKTTLRIETDNTLSKLVQAQALPTENYQINITI-DGFTLQVVERPPHFNDP 650
QY 507 KKPYPILPTPYGGGAGQEVTRKQWALNFKAYVASDSELEYVTWVDRGTGFGKGRFSAV 566
Db 651 RKPYPVLFHLYGGPGSQTVDRREN-VDFQSYVA--ASLGIYVTVVDRGTGFGFGRAARCI 707
QY 567 TRQGLLEAEQDIYAAQQAANIPIWADHIGWGSFGVGLTSKVLKDSG-AFTLGVI 625
Db 708 RGNIGHYEAIDQATAKNAQKPYDSESAWGWSYGGFWMTLKTLEQDAGETFGYGMV 767
QY 626 APVSDWRFYDSMYTERYMKTLSTNEGYETSAVRKTDGFKNVGEGFLIQHGTGDDNVHFO 685
Db 768 APVSDWRFYDSMYTERYMKTLSTNEGYETSAVRKTDGFKNVGEGFLIQHGTGDDNVHFO 826
QY 686 NSALVDLLMGDGVSPKHLHSQWFTSDHSGISVHGGGVFLYKQ 728
Db 827 NTLTLDLKL--DLASQVQYDVHFYPDSHSHSIFPNHAHTMYIER 867

RESULT 12
Q7SHU8 NEUCR
ID Q7SHU8 NEUCR PRELIMINARY; PRT; 895 AA.
AC Q7SHU8;
DT 01-MAR-2004 (T:EMBLrel. 26, Created)
DT 01-MAR-2004 (T:EMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (T:EMBLrel. 26, Last annotation update)
DE Hypothetical protein.
GN Name=NCU02515.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrenikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysseles M., Mauceli E., Biele C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbold D.J., Freitag M.,
```

```
RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0 (2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DDAJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000002; EAA36410.1; -; Genomic_DNA.
DR HSSP; P27487; 1NU6.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004274; F:dipeptidyl-peptidase IV activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR001375; Peptidase S9.
DR InterPro; IPR002469; Peptidase S9B.
DR InterPro; IPR000379; Ser. esters.
DR Pfam; PF00930; DPPIV_N; I.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.
DR PROSITE; PS00708; PRO ENDOPEP SER; UNKNOWN 1.
KW Hydrolase; Hypothetical protein; Transmembrane.
SQ SEQUENCE 895 AA; 100692 MW; F036509E1E3D2B6D CRC64;

Query Match 31.7%; Score 1276.5; DB 2; Length 895;
Best Local Similarity 38.1%; Pred. No. 9.3e-73;
Matches 301; Conservative 134; Mismatches 281; Indels 73; Gaps 30;

QY 7 PHAPTEGS-KELTNETVVVKQAIPTSRSVOWLSG--AEDG-----SYVAAED 53
Db 118 POATVTHSGGKVTLDQ-VLNNEWRKSHSISWAGVNGEDGLLLEKEGANKDYLVVEDV 176
QY 54 GSLLTIENIVTNSRTLI PADKIPTGKEAFN--YWIHP-----DLSSVLASNHTKQVR 104
Db 177 RAQNPSSEVZASKTLL-KDKL---PEFANKTYW--PTVTVPSSDLKKVLLATVDQNNWR 230
QY 105 HSFFADYVQDVESLKSVPFLMP-DOEGDIQYQWSPVGNITAFVRENLDLV--WDNGTVT 161
Db 231 HSYAVYVWIFDVETQAEPLVPYDADARLQLASWSPTSDAIVVYTRDNNFLRLKDSDKIV 290
QY 162 RITDGGPDMFHGVPDWIYEELGDRVALWFSPPGEVLAFLSNFTGVPYTYVQYMDN-- 219
Db 291 QITRDGSADVFNQVPDWIYEELVSGVATWSEDGNYVAFLTNETIGVPEYFIQFVSR 350
QY 220 -DNOELIAPA---YPWELKIRYPKVSQTNPVTLSLNLNIAKSKYKQAPIDA-FESTDLII 274
Db 351 PSGEPEKPGEEYVPEVRQIKYKAGAHNPVDLKFYDVKRGDVFSDISGRPADDDRLIT 410
QY 275 EVAWLTDTHTTVAAKAFNRVQDQKVAVDATSN-----KATVISDRDGTGWLNDLLSM 329
Db 411 EVIWI---AGQVLLIKETNRVSDVMRVVLVDVSGSRGTGKAVRTVDVNDIDG--GWFEISHT 465
QY 330 KYIGPIKPSD--KDYYIDISHSQWAHL-YLFPVSGGEPILTKGDWEVTSI-LSISOEROL 385
Db 466 KFI-PADPANGRPDDGVDYVTHINNGDHLAYFTPLDNPPEPIMLTSGEWEVDAPTAVDLQ 524
QY 386 ROLVYVLSQHHSTERHLYSVSYSTFAVTPLVDDTVAAYWSASFANSGYIILTYGGPDPV 445
Db 525 RNLVYFVSTKRESSIQRHVQVKLTGEDMPTVDTTSKEGYAIFSFGAGYALVSQYQGPNI 584
QY 446 PYQELYTT--NSTKPLRTITDQNAKVLQIKDYALPNITYFELPLPSGETLNVQRLPPG 503
Db 585 PWQKVLSTSPNPKYEHVVEENKDLAEAAKKHELPINIYGTINV-DGVELNVIERRPPHF 643
QY 504 SPDKKYPIILPTPYGGGAGQEVTRKQWALNFKAYVASDSELEYVTWVDRGTGFGKGRFS 563
Db 644 DKMKKYEVLFVQYSGSPVSTVKKTF-AVDFQSFVAAG--LGYICVTVDGRTGFGIGRKNR 700
QY 564 SAVTRQGLLEAEQDIYAAQQAANIPIWADHIGWGSFGVGLTSKVLKDSG-AFTLG 622
Db 701 VITRGNLGTWESHQDIYAAKHWAKQYIDEDRLAIWGSYGGVWYTLKTLSDAGQTFYK 760
QY 623 VITAPVSDWRFYDSMYTERYMKTLSTNEGYETSAVRKTDGF-KNVEGGFLIQHGTGDDN 681
```


RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokytas T., Lokytas Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Maraballa R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., McGhee T., Meldrim J., Meneus L.,
RA Mesirov J., Mihailev A., Mihova G., Mikkelson T., Mieng V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., Neves C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizari M., Norbu C.,
RA Norbu N., O'donnell P., Okaewo O., O'leary S., Omotosho B.,
RA O'Neill K., Oman S., Parker S., Perrin D., Phunkhang P., Pignani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutnan M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA Stevenson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tensing P., Tesfaye S., Theodore J., Thouloung Y., Topham K.,
RA Towey S., Tsamila T., Tsomo N., Vallee D., Vassiliev H.,
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.;
RT "The genome sequence of Magnaporthe grisea";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-70-15;
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN-70-15;
RA Zhu H., Blackmon B.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACU01000945; EAA53468.1; -; Genomic_DNA.
KW Hypoetical protein.
SQ SEQUENCE . 938 AA; 105280 MW; 394BD6652B2C8283 CRC64;

Query Match
Best Local Similarity 31.0%; Score 1248.5; DB 2; Length 938;
Matches 301; Conservative 116; Mismatches 304; Indels 75; Gaps 26;

QY 4 PRKHAPTGEGSKRLTFNVTVKQALTPTSRSVQLSG--AEDGSVYAAEDG-----54
Db 135 PEKTKAAALTSKKITMDQ-VRSGTWYAAASHDITWIAAGPKGSDGLMIEVGADGKDYIVIE 193
QY 55 -----SLTIENIVTNESTRILIPADK-----IPTKEAFNYIHPDLSSVLWAS 97
Db 194 DGRAENPDSSAAEVDVTSRRITLMKNGSFRVQGRWHIFSGAEPSP-----DQCKVLLRT 248
QY 98 NHTKQYRHFADYVYQDVESLKSVPIMP-DOEGDIQYAGWSPGVNNTIAFVRENDLYVWD 156
Db 249 EQKSNWRHSHGHTACVYNDVATETAEELIPASPDARVQNAQWSPSTDAIVFTRENNLYRT 308
QY 157 NGT--VTRITDDGPPNFGVDPDWIYEEELIDGRYALWSPDGEYLAYSFNETGYPTVT 214
Db 309 IGSKKVQITKDGADLFGVDPDWYEEVEIAGASTTWSEDKGYAFRLTNETGVPAFP 368
QY 215 VQYMDNQETA-----PAYMELKIRYKPVQNTPTVLSLNLINAKSEVKQAPID-APE 267
Db 369 VQYFMSRSPGSPKAGEETPEVRDIKPRSGSHNPVMDVQFYDIARGDVFTVDINGGFA 428
QY 268 STLIIIGEAWLTDHTTAAKAFNRVQDQKVVAVD----TASNKATV-ISRDRDGTGW 322
Db 429 DEDRLITVJLVN---ACQVIVKETNRVSDIMRVVLIDVKRGTGKTRTVDVGVKIDG--GW 483
QY 323 LDNLLSMKYI--GPIKPSDKAYYIIDSHSWAHL-YLFPVSGGSPILPTKG-DHEVTS 378
Db 484 FEISQNTKVI PADPKGRDQDQ-YIDTVIHNDGDHLYAFSPDPNPDPFIMLTGSGWEVVD 542

QY 379 ILS-IDQRQLVYVYLSTQHHSTERHLISVSYSTFAVTLVDVDTVAAYHSASFSANSYYI 437
Db 543 APSADVLENNLVYFIATKEGSTORHVSVQLDGNKMKSFDTTEAKGYDYVDSFGAGYAL 602
QY 438 LTYGGDPVPYQELVTT--NSTKPLRTITDNKLVLEQIKDYALPNITYFELPLPSETLNV 495
Db 603 LSYGPKPKIPWQKVSTPANNORYEVLIEENKELASARKYDLPLNYGLTANV-EGVELNY 661
QY 496 MORLPPGSPDKKYPILFTPYGGGAQVETKRWQALNFKAVYASDSLEYVYVTTVDNRGT 555
Db 662 LERRPPQSEKKYPVLPHQYSGPGSQSVSQKF-SVDFQSYVA--ASLGIYVVTDFGRGT 718
QY 556 GFKGRKFRSATVTRQGLLEADSDIYAAQAAANIPIWDADHICIGWCSGGYLTSTKVLKXD 615
Db 719 GFPGKRNVLVRSRLEGEIEAQDQIAAAKHWSLGYVDPFSRIAIGWMSYGGFOTLKTLEAD 778
QY 616 SG-AFTLGIVITAPVSDWRFYDSMYTERYMKLTSTNEEGVETSAVRKTDGFKNVEGGLIQ 674
Db 779 AGRTFSYGMVAPVTDWRFYDSIYTERYMLTFPQONEDGYTKSAVHNVSALAS-NKRFLLM 837
QY 675 HGTGDDNVHFQNSAALVDLLMGDGVSPKLSHSQWFTSDSHGISYHGGVFLYKQARKLY 734
Db 838 HGASDDNVHFQNSLTLKLDMGAVENYDVHV--FPDSHDSIFFHNANKIYDKLENLVLV 895
QY 735 QEKNRQTQVLMHQWTK 750
Db 896 NAFN-----GEWLK 904

RESULT 15
Q6CFX2 YARLI PRELIMINARY; PRT; 868 AA.
AC Q6CFX2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to tr|Q96VT7|Aspergillus niger Dipeptidyl aminopeptidase type IV.
GN OrderedLocusNames=YALI0B02838g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxId=4952;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CLIB 122 / E 150;
RX PubMed=1529592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E., V.,
RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykaaten C.,
RA Boiarane A., Boyer J., Cattolico L., Confaniolero F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Niclaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zentou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudouet B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
DR EMBL; CB382128; CAG82658.1; -; Genomic_DNA.
DR GO; GO:0016020; C.membrane; IEA.
DR GO; GO:0004177; F.aminopeptidase activity; IEA.
DR GO; GO:0003824; F.catalytic activity; IEA.
DR GO; GO:0004274; F.dipeptidyl-peptidase IV activity; IEA.
DR GO; GO:0006508; P.proteolysis and peptidolysis; IEA.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR002469; Peptidase_S9B.
DR InterPro; IPR000379; Ser_estrs.
DR Pfam; PF00930; DPPIV_N; 1.

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2006, 03:12:40 ; Search time 27 Seconds
(without alignments)
2311.857 Million cell updates/sec

Title: US-10-723-807-2_COPY_17_771

Perfect score: 4033

Sequence: 1 LDVPRKPHAPTGGSKRLTF.....EKNRQTQLMHQWKDLE 755

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5 COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/6 COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/H COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/RE COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfilee1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4033	100.0	771	2	US-09-079-592-2
2	3886	96.4	771	2	US-09-462-284-2
3	1054	26.1	818	2	US-09-462-845-3
4	1054	26.1	818	2	US-10-402-312-3
5	1054	26.1	818	2	US-10-401-437-3
6	1054	26.1	818	2	US-10-402-067-3
7	1054	26.1	818	2	US-10-401-436-3
8	987.5	24.5	755	4	PCT-US93-07923-3
9	987.5	24.5	759	4	PCT-US93-07923-2
10	987.5	24.5	766	1	US-08-230-491A-3
11	987.5	24.5	766	1	US-08-619-280A-3
12	987.5	24.5	766	1	US-08-940-391-3
13	987.5	24.5	766	2	US-09-794-236-1
14	987.5	24.5	766	2	US-09-265-606-3
15	987.5	24.5	775	2	US-09-949-016-10450
16	981.5	24.3	766	2	US-10-002-593-6
17	981.5	24.3	766	2	US-09-949-016-6146
18	981.5	24.3	766	2	US-09-518-550-27
19	981.5	24.3	766	2	US-10-423-714-6
20	977	24.2	760	1	US-08-230-491A-2
21	977	24.2	760	1	US-08-619-280A-2
22	977	24.2	760	1	US-08-940-391-2
23	977	24.2	760	2	US-09-265-606-2
24	972	24.1	761	2	US-09-518-550-26
25	902.5	22.4	931	2	US-09-079-592-11
26	869.5	21.6	723	2	US-09-518-550-29
27	865.5	21.5	676	2	US-09-518-550-42

28	811.5	20.1	710	2	US-09-518-550-28	Sequence 28, Appl
29	803.5	19.9	743	2	US-10-363-937-4	Sequence 4, Appl1
30	803.5	19.9	796	2	US-09-976-674-5	Sequence 5, Appl1
31	793.5	19.7	706	2	US-09-976-674-41	Sequence 41, Appl
32	770.5	19.1	494	2	US-09-248-796A-19472	Sequence 19472, A
33	770.5	18.7	691	2	US-09-976-674-43	Sequence 43, Appl
34	638.5	15.8	526	2	US-09-248-796A-19924	Sequence 19924, A
35	623	15.4	732	2	US-09-518-550-30	Sequence 30, Appl
36	613.5	15.2	593	4	PCT-US93-07923-11	Sequence 11, Appl
37	582.5	14.4	882	2	US-09-976-674-1	Sequence 1, Appl1
38	582.5	14.4	882	2	US-10-070-464-1	Sequence 1, Appl1
39	542.5	13.5	547	2	US-09-949-016-8330	Sequence 8330, Ap
40	506.5	12.6	981	2	US-09-902-540-16812	Sequence 16812, A
41	458	11.4	863	2	US-09-976-674-3	Sequence 3, Appl1
42	458	11.4	892	2	US-09-976-674-23	Sequence 23, Appl
43	458	11.4	892	2	US-09-976-674-27	Sequence 27, Appl
44	456	11.3	323	2	US-09-270-767-45296	Sequence 45296, A
45	421.5	10.5	879	2	US-09-976-674-33	Sequence 33, Appl

ALIGNMENTS

RESULT 1

US-09-079-592-2
; Sequence 2, Application US/09079592B
; Patent No. 6664092
; GENERAL INFORMATION:
; APPLICANT: Alexander Blinkovsky
; APPLICANT: Kimberly Brown
; APPLICANT: Michael W. Rey
; APPLICANT: Alan Klotz
; APPLICANT: Tony Byun
; TITLE OF INVENTION: Polypeptides Having Dipeptidyl
; TITLE OF INVENTION: Aminopeptidase Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 5254.200-US
; CURRENT APPLICATION NUMBER: US/09/079,592B
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 08/857,884
; PRIOR FILING DATE: 1997-05-15
; PRIOR APPLICATION NUMBER: 60/062,892
; PRIOR FILING DATE: 1997-10-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 771
; TYPE: PRT
; ORGANISM: Aspergillus oryzae
US-09-079-592-2

Query Match	100.0%	Score 4033;	DB 2;	Length 771;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches	755;	Conservative	0;	Mismatches
			0;	Indels
			0;	Gaps
Qy	1	LDVPRKPHAPTGGSKRLTFNETVVKQAITPTTSRSVQWLSGAEDGSYVYAAEDGSLTIEN	60	
Db	17	LDVPRKPHAPTGGSKRLTFNETVVKQAITPTTSRSVQWLSGAEDGSYVYAAEDGSLTIEN	76	
Qy	61	IVTNESRTLIPADKIPGTGKEAFNYWIHPDLSSVLWASNHTKQYRHSFFADYYQDVESLK	120	
Db	77	IVTNESRTLIPADKIPGTGKEAFNYWIHPDLSSVLWASNHTKQYRHSFFADYYQDVESLK	136	
Qy	121	SVPLMPDQEGDIOYAQWSPVGNITAFVRENDLYVWDNGTVTTRIITDDGCPDMFHGVPDWIY	180	
Db	137	SVPLMPDQEGDIOYAQWSPVGNITAFVRENDLYVWDNGTVTTRIITDDGCPDMFHGVPDWIY	196	
Qy	181	EEETILGDRYALWFSPDGEYLAYSFNETGVPTTYVQYMDNQEIAPAYPWEKIRYPKVS	240	
Db	197	EEETILGDRYALWFSPDGEYLAYSFNETGVPTTYVQYMDNQEIAPAYPWEKIRYPKVS	256	
Qy	241	QTNPTVTLSLNTASKEVKQAPIDAFESTDLIIIGEVAWLTDTHTTTAAKAFNRVQDOQKV	300	
Db	257	QTNPTVTLSLNTASKEVKQAPIDAFESTDLIIIGEVAWLTDTHTTTAAKAFNRVQDOQKV	316	

```

QY 301 VAVDTASNAKATVISDRDGTGWLNLNLSMKYIGIPKPSDKDAYYIDISDHSGWAHLYLFP 360
DB 317 VAVDTASNAKATVISDRDGTGWLNLNLSMKYIGIPKPSDKDAYYIDISDHSGWAHLYLFP 376
QY 361 VSGGEPIPLTKGDMWEVTSILSIDQERQLVYVLSLSTQHHSSTERHLYSVSYSTFAVTPPLVDDT 420
DB 377 VSGGEPIPLTKGDMWEVTSILSIDQERQLVYVLSLSTQHHSSTERHLYSVSYSTFAVTPPLVDDT 436
QY 421 VAAVWSASFNSANGYIILTYGGPDVPYQELVTTNSTKPLRTITONAKVLEQIKDYALPNI 480
DB 437 VAAVWSASFNSANGYIILTYGGPDVPYQELVTTNSTKPLRTITONAKVLEQIKDYALPNI 496
QY 481 TYFELPLPSGETLNVMQRLPPGSPDKKYPILFTPYGGPGAQEVTKRQALNFKAYVASD 540
DB 497 TYFELPLPSGETLNVMQRLPPGSPDKKYPILFTPYGGPGAQEVTKRQALNFKAYVASD 556
QY 541 SELEYVTVTDNRGTGFKGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPTWIDADHIGTW 600
DB 557 SELEYVTVTDNRGTGFKGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPTWIDADHIGTW 616
QY 601 WSGFGYLTSKVLKXDSGFTLGVIITAPVSDRWFYDSMYTERYMKTLSTNEEGYETSARK 660
DB 617 WSGFGYLTSKVLKXDSGFTLGVIITAPVSDRWFYDSMYTERYMKTLSTNEEGYETSARK 676
QY 661 TDGFKNVEGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQSOWFTDSHGHSYHG 720
DB 677 TDGFKNVEGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQSOWFTDSHGHSYHG 736
QY 721 GGVFLYKQLARKLYQEKNRQTQVLMHQTCKDLEE 755
DB 737 GGVFLYKQLARKLYQEKNRQTQVLMHQTCKDLEE 771

RESULT 2
US-09-462-284-2
; Sequence 2, Application US/09462284
; Patent No. 6309868
; GENERAL INFORMATION:
; APPLICANT: Nestec S.A.
; APPLICANT: Monod, Michel
; APPLICANT: Doumas, Agnes
; APPLICANT: Affolter, Micheal
; APPLICANT: Van Den Broek, Peter
; TITLE OF INVENTION: CLONING OF THE
; TITLE OF INVENTION: PROLYL-DIPEPTIDYL-PEPTIDASE FROM
; FILE REFERENCE: 8265-298
; CURRENT APPLICATION NUMBER: US/09/462,284
; CURRENT FILING DATE: 2000-01-03
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FactSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 771
; TYPE: PRT
; ORGANISM: Fungus
US-09-462-284-2

Query Match 96.4%; Score 3886; DB 2; Length 771;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 733; Conservative 1; Mismatches 19; Indels 4; Gaps 2;

QY 1 LDVPRKPHAPTGGSKRLTFNETVVKQAITPTSRSVQWLSGAEDGSGYVYAAEDGSLTIEN 60
DB 17 LDVPRKPHAPTGGSKRLTFNETVVKQAITPTSRSVQWLSGAEDGS--LRVGRGRQSH 74
QY 61 IVTNEST--LIPADKIPTKGFAFNWYIHPDLSVLWASNHTKQYRHSFFADYVVDVES 118
DB 75 REHRQVTHAHADKIPTKGFAFNWYIHPDLSVLWASNHTKQYRHSFFADYVVDVES 134
QY 119 LKSVPLMPDQEGDIOQAQWSFVGNITAFVRENDLYVWDNGTVTRITDDGPDMPHFQVPM 178
DB 135 LKSVPLMPDQEGDIOQAQWSFVGNITAFVRENDLYVWDNGTVTRITDDGPDMPHFQVPM 194

```

```

QY 179 IYEEBILGDRYALWFSPDGEYLAYLSFNETGVPTTYVQYTMNQELIAPVPWELKIRYPK 238
DB 195 IYEEBILGDRYALWFSPDGEYLAYLSFNETGVPTTYVQYTMNQELIAPVPWELKIRYPK 254
QY 239 VSGTNPVTLSLNLNLSIAKQVQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDQ 298
DB 255 VSGTNPVTLSLNLNLSIAKQVQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDQ 314
QY 299 KVAVDTASNAKATVISDRDGTGWLNLNLSMKYIGIPKPSDKDAYYIDISDHSGWAHLYL 358
DB 315 KVAVDTASNAKATVISDRDGTGWLNLNLSMKYIGIPKPSDKDAYYIDISDHSGWAHLYL 374
QY 359 FPVSGGEPIPLTKGDMWEVTSILSIDQERQLVYVLSLSTQHHSSTERHLYSVSYSTFAVTP 418
DB 375 FPVSGGEPIPLTKGDMWEVTSILSIDQERQLVYVLSLSTQHHSSTERHLYSVSYSTFAVTP 434
QY 419 DTVAAYWSASFNSANGYIILTYGGPDVPYQELVTTNSTKPLRTITONAKVLEQIKDYALP 478
DB 435 DTVAAYWSASFNSANGYIILTYGGPDVPYQELVTTNSTKPLRTITONAKVLEQIKDYALP 494
QY 479 NITVFELPLPSGETLNVMQRLPPGSPDKKYPILFTPYGGPGAQEVTKRQALNFKAYVA 538
DB 495 NITVFELPLPSGETLNVMQRLPPGSPDKKYPILFTPYGGPGAQEVTKRQALNFKAYVA 554
QY 539 SDSELEYVTVTDNRGTGFKGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPTWIDADHIGI 598
DB 555 SDSELEYVTVTDNRGTGFKGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPTWIDADHIGI 614
QY 599 WGSFGGYLTSKVLKXDSGFTLGVIITAPVSDRWFYDSMYTERYMKTLSTNEEGYETSARV 658
DB 615 WGSFGGYLTSKVLKXDSGFTLGVIITAPVSDRWFYDSMYTERYMKTLSTNEEGYETSARV 674
QY 659 RKTGDFKNVEGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQSOWFTDSHGHSY 718
DB 675 RKTGDFKNVEGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQSOWFTDSHGHSY 734
QY 719 HGGVFLYKQLARKLYQEKNRQTQVLMHQTCKDLEE 755
DB 735 HGGVFLYKQLARKLYQEKNRQTQVLMHQTCKDLEE 771

RESULT 3
US-09-462-845-3
; Sequence 3, Application US/09462845
; Patent No. 6723550
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/09/462,845
; CURRENT FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-09-462-845-3

Query Match 26.1%; Score 1054; DB 2; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-81;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

QY 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRSVQWLS-----GAEDGSYV 48
DB 48 IPHSNTPDYQEPNSNYTNDGKLVKSFS--VVRNNTFQPKYHELOWISDNKIESNDLGLYV 106
QY 49 YAAEDGSLTIENIVNESRTLLIPADKIPTKGFAFNWYI-----PDLSSVLW 95

```

Db 107 TFND-SYVVKSYDDSYNSVLEGGT-----FIHQNLTVESITASDLKRLI 156
Qy 96 ASNHTKQYRHSFADYVQVDSKSLKVPMLPDQEGDIQYAOQSPVGNNTIAFVRENDLYW 155
Db 157 RTNSVQNRHSFSGSYFYVD---KSSSFEEIGNEVALAINSPNSNDIAYQDNNIYIY 212
Qy 156 ---DNGTVTRITDDGDPDMFHGVDWYEEIILGDRYALWFSFDGSLAYLSFNETGVPT 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAWSPGTGYLAFKIDSEVGE 272
Qy 213 YTVQYMDNQEIAPAYPWELKIRYKVSQTNPTVTLNLNIAASKYKQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIYPMRSIKYKSGTNPNAELWYVSMKDGTSFHPRISGNKKGSL 330
Qy 272 IIGEVAVLTDTHTTVAAKAFNRVQDQKVVAVDTASNKATVISDRDGTGWLNDLLSMKY 331
Db 331 LITEVTWVNGN--VLVKTDRSSDILTFLDITIAKTSNVVRNESSNGGWEITHNTLF 388
Qy 332 IGIKPSDKDAY-YIDISHSQWALYLPVPSGGEPI-PLTKGDWEVTS-ILSIDQEROL 388
Db 389 IPANETFDPRHNGYVDILPFGYNHLAYFENSSSHYKTLTEGKWEVNGPLAFDSMENR 448
Qy 389 VYVLSQTHSTHERHLYSVSY-STFAVTPLVDDTVAAYWSASFANSYVILTYGGPDVY 447
Db 449 LYFISTRKSTERHYYIDLRSPNEIEVTDSEDGVYDVSPSSGRRFGLTYGPKVPY 508
Qy 448 QELYTNTST-----KPLRTITDNKVLQIKDYALPNITYFELPLPSGE----491
Db 509 QKIVDFHSRKAECDCGNVLGSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDFGKD 568
Qy 492 -TLNVQRLPPGSP--DKKYPILFTPYGGGAEVTKRWQALNFKAYVASDSELYVTW 548
Db 686 LKTLKDGGRHFKYGSVAPVTDWRFYDSYTYRYMHTPOENFDGYVSSVHNVTAQA 745
Qy 688 EGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSQWFTDSHGISHYGGVFLYK 727
Db 746 -NRFLMHGTGDDNVHFQNSLKFLLDLGLNGVENYDVHV--FPDSHSIRYHNAVIVFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 4

US-10-402-312-3
; Sequence 3, Application US/10402312
; Patent No. 683261
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/10/402,312
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-402-312-3

Query Match 26.1%; Score 1054; DB 2; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-81;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQMLJ-----GAEDGSYV 48
Db 48 IPHNSNTPDYQBPNSNYTNDGKLVKVSFS-VVRNNTFPQPKYHLOWISDNKIESNDLGLYV 106
Qy 49 YAAEGSLTIENIVNESRTLLPADKIPGKEAFNYIHH-----PDLSSVLW 95
Db 107 TFND-SYVVKSYDDSYNSVLEGGT-----FIHQNLTVESITASDLKRLI 156
Qy 96 ASNHTKQYRHSFADYVQVDSKSLKVPMLPDQEGDIQYAOQSPVGNNTIAFVRENDLYW 155
Db 157 RTNSVQNRHSFSGSYFYVD---KSSSFEEIGNEVALAINSPNSNDIAYQDNNIYIY 212
Qy 156 ---DNGTVTRITDDGDPDMFHGVDWYEEIILGDRYALWFSFDGSLAYLSFNETGVPT 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAWSPGTGYLAFKIDSEVGE 272
Qy 213 YTVQYMDNQEIAPAYPWELKIRYKVSQTNPTVTLNLNIAASKYKQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIYPMRSIKYKSGTNPNAELWYVSMKDGTSFHPRISGNKKGSL 330
Qy 272 IIGEVAVLTDTHTTVAAKAFNRVQDQKVVAVDTASNKATVISDRDGTGWLNDLLSMKY 331
Db 331 LITEVTWVNGN--VLVKTDRSSDILTFLDITIAKTSNVVRNESSNGGWEITHNTLF 388
Qy 332 IGIKPSDKDAY-YIDISHSQWALYLPVPSGGEPI-PLTKGDWEVTS-ILSIDQEROL 388
Db 389 IPANETFDPRHNGYVDILPFGYNHLAYFENSSSHYKTLTEGKWEVNGPLAFDSMENR 448
Qy 389 VYVLSQTHSTHERHLYSVSY-STFAVTPLVDDTVAAYWSASFANSYVILTYGGPDVY 447
Db 449 LYFISTRKSTERHYYIDLRSPNEIEVTDSEDGVYDVSPSSGRRFGLTYGPKVPY 508
Qy 448 QELYTNTST-----KPLRTITDNKVLQIKDYALPNITYFELPLPSGE----491
Db 509 QKIVDFHSRKAECDCGNVLGSLYHLEKNEVLTKILEDYAVPRKSFRELNLGKDFGKD 568
Qy 492 -TLNVQRLPPGSP--DKKYPILFTPYGGGAEVTKRWQALNFKAYVASDSELYVTW 548
Db 569 ILVNSYEILPNDFDETLSDHYPVFFPAYGPNSSQVVKTF-SVGFNEVVA--SQLNAIVV 625
Qy 549 TVDNRGTFGKGRFSAVTRQLGLEAEDQIYAAQAAANIPWIDADHIGIWSFGYLT 608
Db 626 VVDGRGTGFGQDFRSLVRDLGDEARQISAAVLSYGLTFFVDPQKISLFGWSYGYLT 685
Qy 609 SKVLEKDSGA-FTLGVTITAPVSDWRFYDSMYTERYMKTLSTNEEGYETSAVRKTDGKPV 667
Db 686 LKTLKDGGRHFKYGSVAPVTDWRFYDSYTYRYMHTPOENFDGYVSSVHNVTAQA 745
Qy 688 EGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSQWFTDSHGISHYGGVFLYK 727
Db 746 -NRFLMHGTGDDNVHFQNSLKFLLDLGLNGVENYDVHV--FPDSHSIRYHNAVIVFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 5

US-10-401-437-3
; Sequence 3, Application US/10401437
; Patent No. 684940
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/10/401,437
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: PCT/US98/14647

QY 668 EGGFLIOHGTGDDNVHFNQSAALVDLLMGDGVSPKLSHQWFTDSDHGISYHGGGVFLYK 727
DB 746 -NRFLMHGTGDDNVHFNQSLKFLDLDDLVGVNDVHV--FPDSHRSIRYHNANVIFD 802
QY 728 QL 729
DB 803 KL 804
RESULT 7
US-10-401-436-3
; Sequence 3, Application US/10401436
; Patent No. 6911333
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Processes From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/10/401,436
; PRIOR FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-401-436-3
Query Match 26.1%; Score 1054; DB 2; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-81;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;
QY 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
DB 48 IPHNSNTPDQENSNYNDGKLKVSFS-VVRNNTFQPKYHQLQWISDNKIESNDLGLYV 106
QY 49 YAAEDGSLTIENIVNESLTPADKIPITGKEAFNWIH-----PDLSSVLW 95
DB 107 TFWND-SYVVKSYVDYSNLSLEGT-----FIHQNLVATESITASPDRLKLLI 156
QY 96 ASNHTQYRHSFADYVVDVSVESLKSVPMPQDEQDIQVAKSPVGNITAFVRENDLYW 155
DB 157 RNSVQNRHSTPGSFYVD-----KSSSFEIEGNEVALAIWSPNSNDIAYQDNNIYI 212
QY 156 ---DNGTVTRITDDGPDMPHGVDPDIYBEEIIGDRYALWFSFGDEYLAFLYSFNETGVPT 212
DB 213 SAISKKTIRAVNDGSSFLFNGKPDVYEEVPEDDKAAWSPDGYLAFLKIDSEVGE 272
QY 213 YTVQYTMNOETAPAYPMELKIRYKVSQTNPTVTLNLNIASKEVKQAPIDAFEST-DL 271
DB 273 FIPIYVQDEK--DIYPMRSIKYKSGTNPNAELWVYSMDGTSFHPRISSGNKKGSL 330
QY 272 IIGEVAVLTDTHTTAAKAFNRVQDQKVAVDTASNKATVLSDRDGTGDLNLLSMKY 331
DB 331 LITEVTWVNGN--VLVKTTRSSDILTFLDITIAKTSNVVNESNGGWEIFHTLTP 388
QY 332 IGPIKPSDKDAY-YIDISHSQWALYLFVPSGGEPI-PLTKGDWVTS-ILSIDQERQL 388
DB 389 IPANETFDPHNGYVDILPIGGYNHLAYFENSNSHYKTLTEGKEVNVNGPLAFDSMENR 448
QY 389 VYVLSQTHSTHERHLYSVSY-STFVATPLVDDTVAAVWSASFSAFGSYILTYGGDPVY 447
DB 449 LYPISTRKSTERRHYVYIDLRSPNEIIEVTDSEDGVYDVSFSGRRRFGLLTYKGPVY 508
QY 448 QELYTNTST-----KPLRTITDNAKVLQIKDVALPNITYFELPLPSGS---- 491
DB 509 QKIVDFHRSKAEKCDKGNVLGKSLYHLEKNNEVLTKEILDYAVPRKSFRELNLKDFGKD 568
QY 492 -TLNVNQRILPPGFS--DKKYPIFLTPYGGPGAQEVTKRWQALNFKAYVASDSELYVTW 548

DB 569 ILVNSYEILPNDFDTLSHDYVPVFFPFAVGPNSSQVVKTF-SVGNEVVA--SQLNAIVV 625
QY 549 TVDNRTGFKGRKFRSAVTROQLGLLEADQIYAAQAAANIPWIDADHIGIWMGSGGYLT 608
DB 626 VVDGRTGFKGQDFSLVDRDLGDEYARDQISAAALYGLSLTVPDPQKISLFCWSYGGYLT 685
QY 609 SKVLEKDSGA-FTLGVIITAPVSDWRFPYDSMYTERMKTLSTNBEYGETSAVRKTDGFKNV 667
DB 686 LKLTLEKGGRHFKYGMVAPVTDWRFPYDSVYTERYMTFPQENFDGYVSSVHNVTALQA 745
QY 668 EGGFLIOHGTGDDNVHFNQSAALVDLLMGDGVSPKLSHQWFTDSDHGISYHGGGVFLYK 727
DB 746 -NRFLMHGTGDDNVHFNQSLKFLDLDDLVGVNDVHV--FPDSHRSIRYHNANVIFD 802
QY 728 QL 729
DB 803 KL 804
RESULT 8
PCT-US93-07923-3
; Sequence 3, Application PC/TUS9307923
; GENERAL INFORMATION:
; APPLICANT: Morimoto, Chikao
; APPLICANT: Schloesman, Stuart F.
; APPLICANT: Tanaka, Toshiaki
; TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: WordPerfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07923
; FILING DATE: 19930819
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/934,162
; FILING DATE: 21-AUG-1992
; APPLICATION NUMBER: 07/832,211
; FILING DATE: 06-FEB-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Fraser, Janis K.
; REGISTRATION NUMBER: 34,819
; REFERENCE/DOCKET NUMBER: 00530/055002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 755
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
PCT-US93-07923-3
Query Match 24.5%; Score 987.5; DB 4; Length 755;
Best Local Similarity 32.6%; Pred. No. 9.1e-76;
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;
QY 11 TGEGRKLFNETVVKQAITPTSRVQWLSGAEDGSYVVAADGSLTIENIVNESRLLI 70
DB 25 TADSRKTYLTLD-YLKNTRYLKLKYSURWIS---DHEYLKQENILVFNAYCNGSSVFLE 80

```
Qy 71 PADKPTGKEAFNWIHPDLSSVLWASHNTKQYRHSFPADYYVQDVESLKSVPMLPQOE- 129
Db 81 NSTDFEGHSINDYSISPDGQFILLENNYVQWRHSYASY--DIYDLNKRQLITEERI 137
Qy 130 -GDIQYQAWSVPGNTIAFVRENDLYV--WDNGTVTTRITDDGGPD-MFHGVPDWIYEERIL 185
Db 138 PNTQWVTWSPVGHKLAVWNNDIYVKEPNLPSVRIWTGKEDIYNGITDMWYEEVF 197
Qy 186 GDRVALMPSDGEYLAFLSNFNETGPTVTVQYVMDNQBIAPAYPWEKIRPKVSOQNTPT 245
Db 198 SAYSALMWSPNGTFLAYAQFNDTEVPLIEYSFYSD-----SLQYKTVRVVPYKAGAVNPT 254
Qy 246 VTLSELLNT--ASKEVQKAPIDAFESTDLIIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297
Db 255 VKFFVNTDLSLSSVNTNATSIQITAPASMLIGHYLCVDTWATQER--ISLQWLRRIQNY 312
Qy 298 -QKVAVDTASNKATVISDRD-----GTDGMDLNLMSKYIGPIKPS-----DKDAYYID 346
Db 313 VMDICDYDESSGRWNCCLVARQHIEMSTTGW-----VGRFRPSEPHFTLDGNSFYKI 363
Qy 347 ISDHSGWAHLFLFPVSGGEPILTKGDWEVTSILSIDOERQLVYVLTSTQHHSTE--RHLY 404
Db 364 ISNBEGYRHICVFQIDKKDCTFITKGTWEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 421
Qy 405 SVSVSTFA-VTPL---VDDTVAAYWASFSANSYVILTYGGPDVPYQELVTTNSTKPLR 460
Db 422 KIQLSDYTKVTCLSCELNPERCOYYSVSFSKEAKYVQLRCSGPGGLPLTLHSSVNDKGLR 481
Qy 461 TITDNA---KVLQIK-----DYALPNITYFELPLPSGETLNNVQRLPPGFSPPDKKYP 511
Db 482 VLEDNSALDKMLQNVQMPSSKKLDFILNTEKFWYQM-----ILPPHFDKSKKPYL 531
Qy 512 LFTPYGGGAQEVTKRWQALNFKAYASDSSELYVTTVTDNRGTGFKGRKFRSAVTRQLG 571
Db 532 LLDVYAGCSQKADTVFR-LNWAYTLASTENI--IVASFDRGSGYQGDKIMHAINREL 588
Qy 572 LLEAEDQIYAAQAAANIPIWDADHIGIWMSPGGYLTSTKVLEKSGAFTLVITAPVSDW 631
Db 589 TFEVEDQIEAARQFSKMGFVDNKRITAIWGSYGGVYVTVSMVLGSGSGVFKCGIAPVSRW 648
Qy 632 RPYDSMYTERYM--KTLSTNEBGETSAV-RKTDGFKNVGEGFLIOHGTGDDNVHFQNSA 688
Db 649 EYDYSVYTERYMGFLTPEDNLDHYRNSTVMSRAENFKQVE--YLLIHGTADDNVHFQOSA 706
Qy 689 ALVDILMGDGVSPKLSQWFTSDHGISYHGGGVFLYQLARKLYQ 735
Db 707 QISKALVDVGD---FQAWWYTDHGHASSTAHQHIYTHSHFIKQ 750
```

RESULT 9

PCT-US93-07923-2

Sequence 2. Application PC/TUS9307923

GENERAL INFORMATION:

APPLICANT: Morimoto, Chikao

APPLICANT: Schloesman, Stuart F.

APPLICANT: Tanaka, Toshiaki

TITLE OF INVENTION: HUMAN CD26 AND METHODS FOR USE

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM PS/2 Model 502 or 55SX

OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

SOFTWARE: WordPerfect (Version 5.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/07923

FILING DATE: 19930819

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/934,162

FILING DATE: 21-AUG-1992

APPLICATION NUMBER: 07/832,211

FILING DATE: 06-FEB-1992

ATTORNEY/AGENT INFORMATION:

NAME: Fraser, Janis K.

REGISTRATION NUMBER: 34,819

REFERENCE/DOCKET NUMBER: 00530/055002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 759

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

PCT-US93-07923-2

Query Match 24.5%; Score 987.5; DB 4; Length 759;

Best Local Similarity 32.6%; Pred. No. 9.2e-76;

Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;

```
Qy 11 TGECSKLTFTNETVVKQAITPSTRSVQWLSGAEDGSYVYAAEDGSLTITENIVTNESRTLI 70
Db 29 TADSRKVTTLTD-YLKNTRYRLKLSLRWIS---DHEYLKQENNILVFNAEYGNSSVFLE 84
Qy 71 PADKPTGKEAFNWIHPDLSSVLWASHNTKQYRHSFPADYYVQDVESLKSVPMLPQOE- 129
Db 85 NSTDFEGHSINDYSISPDGQFILLENNYVQWRHSYASY--DIYDLNKRQLITEERI 141
Qy 130 -GDIQYQAWSVPGNTIAFVRENDLYV--WDNGTVTTRITDDGGPD-MFHGVPDWIYEERIL 185
Db 142 PNTQWVTWSPVGHKLAVWNNDIYVKEPNLPSVRIWTGKEDIYNGITDMWYEEVF 201
Qy 186 GDRVALMPSDGEYLAFLSNFNETGPTVTVQYVMDNQBIAPAYPWEKIRPKVSOQNTPT 245
Db 202 SAYSALMWSPNGTFLAYAQFNDTEVPLIEYSFYSD-----SLQYKTVRVVPYKAGAVNPT 258
Qy 246 VTLSELLNT--ASKEVQKAPIDAFESTDLIIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297
Db 255 VKFFVNTDLSLSSVNTNATSIQITAPASMLIGHYLCVDTWATQER--ISLQWLRRIQNY 316
Qy 298 -QKVAVDTASNKATVISDRD-----GTDGMDLNLMSKYIGPIKPS-----DKDAYYID 346
Db 313 VMDICDYDESSGRWNCCLVARQHIEMSTTGW-----VGRFRPSEPHFTLDGNSFYKI 367
Qy 347 ISDHSGWAHLFLFPVSGGEPILTKGDWEVTSILSIDOERQLVYVLTSTQHHSTE--RHLY 404
Db 368 ISNBEGYRHICVFQIDKKDCTFITKGTWEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 425
Qy 405 SVSVSTFA-VTPL---VDDTVAAYWASFSANSYVILTYGGPDVPYQELVTTNSTKPLR 460
Db 426 KIQLSDYTKVTCLSCELNPERCOYYSVSFSKEAKYVQLRCSGPGGLPLTLHSSVNDKGLR 485
Qy 461 TITDNA---KVLQIK-----DYALPNITYFELPLPSGETLNNVQRLPPGFSPPDKKYP 511
Db 486 VLEDNSALDKMLQNVQMPSSKKLDFILNTEKFWYQM-----ILPPHFDKSKKPYL 535
Qy 512 LFTPYGGGAQEVTKRWQALNFKAYASDSSELYVTTVTDNRGTGFKGRKFRSAVTRQLG 571
Db 536 LLDVYAGCSQKADTVFR-LNWAYTLASTENI--IVASFDRGSGYQGDKIMHAINREL 592
Qy 572 LLEAEDQIYAAQAAANIPIWDADHIGIWMSPGGYLTSTKVLEKSGAFTLVITAPVSDW 631
Db 593 TFEVEDQIEAARQFSKMGFVDNKRITAIWGSYGGVYVTVSMVLGSGSGVFKCGIAPVSRW 652
Qy 632 RPYDSMYTERYM--KTLSTNEBGETSAV-RKTDGFKNVGEGFLIOHGTGDDNVHFQNSA 688
Db 653 EYDYSVYTERYMGFLTPEDNLDHYRNSTVMSRAENFKQVE--YLLIHGTADDNVHFQOSA 710
```

```
QY 689 ALVDLLMGDVSPEKLHSONFTSDHGISYHGCGVFLYKQALRKLQ 735
Db 711 QISKALVDVGD---FQAWYITDEDHGIASSTAHQHIYTHMSHFQ 754

RESULT 10
US-08-230-491A-3
; Sequence 3, Application US/08230491A
; Patent No. 5587299
; GENERAL INFORMATION:
; APPLICANT: Rettig, Wolfgang J.; Scanlan, Matthew J.;
; APPLICANT: Garin-Chesa, Pilar; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED NUCLEIC ACID MOLECULE CODING FOR
; TITLE OF INVENTION: FIBROBLAST ACTIVATION PROTEIN ' AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FEFLE & LYNCH
; STREET: 805 THIRD AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE 3.5 inch 1.2 MB STORAGE
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: WORDPERFECT - ASC II
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/230,491A
; FILING DATE: 20-APRIL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5587299man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 330
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 766 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-230-491A-3

Query Match 24.5%; Score 987.5; DB 1; Length 766;
Best Local Similarity 32.6%; Pred. No. 9.3e-76;
Matches 250; Conservative 111; Mismatches 303; Indels 83; Gaps 27;

QY 11 TGGGSKRLTFNVTVVKQAITPTSRVQWLSGAEDEGIVYVAAEDGSLTIENIVTNSRTLI 70
Db 36 TADSRKTYLTLD-YLKNYRLKLYSLRWIS--DHEYLYKQENNILVFNAYGNSSVFLE 91
QY 71 PADKIPGTKEAFNWTHTPDLSSVLMASNHTKQVHSHFFADYVQDVDESLSVPLMPDQE- 129
Db 92 NSTFDFEGHSINDYSPDQGFLLLEYNVYKQWRHSSTASY---DIYDLNKRQLITEERI 148
QY 130 -GDIQVAQSPVGNITAFVRENDLYV--WDNGTVTRITDDGGPD-MFHGVDPDWIYEEEL 185
Db 149 PNTQVTVSPVUGHKLAVVWNNDIYKIEBFLPSYRITWTGKEDIYNGITDVAEEVF 208
QY 186 GDRIALWSPDGBYLAYSFNETGVPTTYVQYMDNQEIAPAYPWLKIRYKVSQTNPT 245
Db 209 SAYSAALWSPNGTFLAYAQENDTEVPLIEYSFYSDE---SLOYPKTVRVPYKAGVNP 265
QY 246 VTLSLNI--ASKEVQKAPIDAPESDTLIIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297
Db 266 VKFFVNTDLSLVSNTATSIQITAPASMLIGDHYLCDVTVATQER--ISLQWLRRIQNS 323
QY 298 -QKVAVAVDTASNKATVISDRD---GTDGWLNDLLSMKYGIPKPS-----DKDAYIID 346

US-08-619-280A-3
; Sequence 3, Application US/08619280A
; Patent No. 5767242
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
; TITLE OF INVENTION: ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280A
; FILING DATE: 18-MARCH-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5767242man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 766 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-619-280A-3
```


QY 572 LLEAEDQIYAAQAAANIPWIDADHIGIWMGSGGGLTSKVLKDSGAFTLGITAPVSDW 631
DB 600 TFEVEDQTEAARQFSKMGFVNDKRIAIWGSYGGVTSWVLGSGGVFKCGIAPVSRW 659
QY 632 RFYDSMYTERYM--KTLSTNEEGYETSAV--RTDGFKNVGGFLIHOHTGDDNVHFQNSA 688
DB 660 EYDSVYTERYMGFLTPEDNLDHYNSVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 717
QY 689 ALVDLLMGDGVSPKELHSGWFTSDHGISYHGGGVFLYKQLARKUYQ 735
DB 718 QISKALVDVGD---FQAMWYTDHGIASSTAHQHIYTHMSHFQK 761

RESULT 13
US-09-794-236-1
; Sequence 1, Application US/09794236
; Patent No. 6337069
; GENERAL INFORMATION:
; APPLICANT: Grouzmann, Eric
; APPLICANT: Lacroix, Jean-Silvain
; APPLICANT: Monod, Michel
; TITLE OF INVENTION: Method of Treating Rhinitis and Sinusitis
; FILE REFERENCE: 81985/276823
; CURRENT APPLICATION NUMBER: US/09/794,236
; CURRENT FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 766
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-794-236-1

Query Match 24.5%; Score 987.5; DB 2; Length 766;
Best Local Similarity 32.6%; Pred. No. 9.3e-76;
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;

QY 11 TGEGRKLTFTNTVVKQAITPTSRVQWLSGAEDSGYVAAEDGSLTIENIVTNSRTLI 70
DB 36 TADSRKTYTLTD-YLKNTRYRLKLSLRWIS---DHEYLKQENNLVFNFAEYGNSSVFLE 91
QY 71 PADKIPTGKEAFNYTHPDLSSVLWASNHTKQYRHSFFADYVQDVESLSKVLPLPDQOB- 129
DB 92 NSTDEFHGSINDYISPDGQFILLIENYVWKWRHSITASY---DIYDLNKRQLITEERI 148
QY 130 -GDIQVAOWSPGNTIAFVRENDLYV--WDNGVTVTRITDDGGPD-MFHGVPMIYEEEL 185
DB 149 PNTQWTVWSPVGHKLAVVWNDIYVKIEPNLPSYRITWTGKEDIYNGITDMVYEEVF 208
QY 186 GDRYALWSPDGEYLAYLSFNETGVPTTYVQYMDNQETAPAYPHELKIRYKVSQTNPT 245
DB 209 SAYSAWSPNGTFLAYAQFNDTEVPLIEYSFYSDE---SLQYPKTVRPYPKAGAVNPT 265
QY 246 VTLSLNI--ASKEVKAQPIDAFESTDLIG-----EVAWLTDTHTVAAKAFNRVQDO- 297
DB 266 VKFFVNTDLSVNTWATSIQTAPASMLIGHYLCVDTWATQER--ISLOWLRQNY 323
QY 298 -QKWAVDATSKATVISDRD-----CTDGLNLLSMKYIGPIKPS-----DKDAYID 346
DB 324 VMDICDYDSSGRWNCVLAHQHIMSTTGW-----VGRFRPSEPHFTLDGNSFYKI 374
QY 347 ISDHGWAHLXYLFPVSGGSPILTKGDWEVTSILSDQERQLVYVLSQHHSTE--RHLY 404
DB 375 ISNEGYRHCYFQIDKDCFTTKGTWIEVIGIEALTSD--YLYYISNEYKMGPGGRNLY 432
QY 405 SVYSSTFA-VTPL---VDTVAAYWASPSANSYVILTYGGPDVPYQELTYTNSKPLR 460
DB 433 KIQLSDYTKVTCLSCENPERCOYYSVFSKEAKYYQLRCSGGLPLTYLHSSVNDKGLR 492
QY 461 TITDNA---KVLBOIK-----DYALPNITYPELPLPSGETLNMORLPFGSPDKKYP 511
DB 493 VLEDNSALDKMLQNVQMPSKLDFIILNETKFWQM-----ILPPHFDKSKKYL 542

QY 512 LFTPYGPGAQEVTKRQWALNFKAYVASDSLEYVTVTVNDRGTGFKGRKFRSAVTRQIG 571
DB 543 LLDVYAGPCSKQADTVFR-LNWAYTLASTENI--IVASFDRGSGYQGDKIMHAINRRLG 599
QY 572 LLEAEDQIYAAQAAANIPWIDADHIGIWMGSGGGLTSKVLKDSGAFTLGITAPVSDW 631
DB 600 TFEVEDQTEAARQFSKMGFVNDKRIAIWGSYGGVTSWVLGSGGVFKCGIAPVSRW 659
QY 632 RFYDSMYTERYM--KTLSTNEEGYETSAV--RTDGFKNVGGFLIHOHTGDDNVHFQNSA 688
DB 660 EYDSVYTERYMGFLTPEDNLDHYNSVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 717
QY 689 ALVDLLMGDGVSPKELHSGWFTSDHGISYHGGGVFLYKQLARKUYQ 735
DB 718 QISKALVDVGD---FQAMWYTDHGIASSTAHQHIYTHMSHFQK 761

RESULT 14
US-09-265-606-3
; Sequence 3, Application US/09265606
; Patent No. 6846910
; GENERAL INFORMATION:
; APPLICANT: Zimmermann, Rainer; Park, John E.;
; APPLICANT: Rettig, Wolfgang; Old, Lloyd J.
; TITLE OF INVENTION: ISOLATED DIMERIC FIBROBLAST ACTIVATION PROTEIN
; TITLE OF INVENTION: ALPHA, AND USES THEREOF
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felle & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 2.0 MB storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/265,606
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/619,280
; FILING DATE: 18-MARCH-1996
; APPLICATION NUMBER: 08/230,491
; FILING DATE: 20-APRIL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 6846910man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 5330.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 766 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-09-265-606-3

Query Match 24.5%; Score 987.5; DB 2; Length 766;
Best Local Similarity 32.6%; Pred. No. 9.3e-76;
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;

QY 11 TGEGRKLTFTNTVVKQAITPTSRVQWLSGAEDSGYVAAEDGSLTIENIVTNSRTLI 70
DB 36 TADSRKTYTLTD-YLKNTRYRLKLSLRWIS---DHEYLKQENNLVFNFAEYGNSSVFLE 91
QY 71 PADKIPTGKEAFNYTHPDLSSVLWASNHTKQYRHSFFADYVQDVESLSKVLPLPDQOB- 129

```
Db 92 NSTDFEGHSINDYSISPDGQFILLLEYNYVQWRHSYASY---DIYDLNKRQLITERI 148
Qy 130 -GDIQYACWSPVGNITIAFVRENDLYV--WDNGTVTRITDDGGPD-MFHVGPDMYIEEIL 185
Db 149 PNNQWVTWSPVGHKLAVWNNDIYVKIEPNLPSYRITWTGKEDIINYNGITDMWYEEVF 208
Qy 186 GDRVALMSPDGEVLAYLSFNETGVPTTYVOYYMNDQEIAPAYPWELKIRPKVQSNPT 245
Db 209 SAYSALMWSNGTFLAYAQFNDTEVPLIEYSFYSD---SLQYPKTVRVVPYKAGAVNPT 265
Qy 246 VTLSLANI--ASKEVKQAPIDAFESTDLIIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297
Db 266 VKFFVWNTDLSLVNATSIQITAPASMLGDHYLDCVWTWATQER--ISLQWLRRIQNY 323
Qy 298 -QKVAVADTASNATVISDRD---GTDGWLNDLLSMKYIGPIKPS-----DKDAYIID 346
Db 324 VMDICDYDESSGRWNCVLAHQHIEMSTIGM-----VGRFRPSPHFTLDGNSFYKI 374
Qy 347 ISDHSGWAHLVLPVPSGGEPIPLTKGDWEVTSILSDOERQLVYLTQHHSTE--RHLY 404
Db 375 ISNEEYRHCYFQIDKKDCTFITKGTWEVIGIEALTSD--YLYYISNEYKMGMPGGRNLY 432
Qy 405 SVSYSTFA-VTPL---VDDTVAAWNSAFSANSQYIILTYGGPDVPYQELVTTNSTKPLR 460
Db 433 KIQLSDYTKVTCLSCELNPERCQYISVSFSKEAKYQLRCSGGPLPLYTLHSSVNDKGLR 492
Qy 461 TITDNA---KYLEQIK-----DYALPNITYFELPLPSGETLNMVQRLPPGSPDKKYP 511
Db 493 VLEDSALDKMLQNVQMPKSLDFIILNETKFWYQ-----ILPHFDKSKKYP 542
Qy 512 LFTPYGGGAQEVTKRQWALNFKAYVASDSLELYVTWTVNDRGTGFGKGRKFRSAVTRQLG 571
Db 543 LLDVYAGPCSQKADTVFR-LNWAYTLASTENI--IVASFDCGRSGYQGDKIMHAINRRLG 599
Qy 572 LLEAEDQIYAAQAAANIPWIDADHIGWGSFGYLTSLKLEKDSGAFLLGVITAPYSDW 631
Db 600 TFEVEDQIEAARQFSKMGFVNDKRIALWGSYGGYVTSWLGSGGVKFCGIAVAPVSRW 659
Qy 632 RFYDSMYTERYM--KTLSTNEEGYETSAV-RKTDGFKNVEGGFLIQHGTGDDNVHFQNSA 688
Db 660 EYDYSVYTERYMGUPTPEDNLDHYRNSTVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 717
Qy 689 ALVDLLMGDGVSPKLSHQWTFSDHGISYHGGGVFLYKQLARKLYQ 735
Db 718 QISKALVDVGD---FQAMWYTDDEHGIASSTAHOHIYTHMSHFQIK 761
```

```
RESULT 15
US-09-949-016-10450
; Sequence 10450, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CLO01307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10450
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10450
```

Query Match 24.5%; Score 987.5; DB 2; Length 775;

```
Best Local Similarity 32.6%; Pred. No. 9.5e-76;
Matches 250; Conservative 131; Mismatches 303; Indels 83; Gaps 27;
Qy 11 TGSCKRLTNETVVKQAIPTSRVQWLSGABDGSVVYAAEDGSLTIENIVTNESSTLI 70
Db 45 TADSRKTYTLTD-YLKNTRYLKLKSLRWIS---DHELYLKQENNILVFNAYGNSVFLP 100
Qy 71 PADKIPITGKAFNWIHPDLSSVLWASNHTKQYRHSFFADYYVQDVESLSKVLMPDQE- 129
Db 101 NSTDFEGHSINDYSISPDGQFILLLEYNYVQWRHSYASY---DIYDLNKRQLITERI 157
Qy 130 -GDIQYACWSPVGNITIAFVRENDLYV--WDNGTVTRITDDGGPD-MFHVGPDMYIEEIL 185
Db 158 PNNQWVTWSPVGHKLAVWNNDIYVKIEPNLPSYRITWTGKEDIINYNGITDMWYEEVF 217
Qy 186 GDRVALMSPDGEVLAYLSFNETGVPTTYVOYYMNDQEIAPAYPWELKIRPKVQSNPT 245
Db 218 SAYSALMWSNGTFLAYAQFNDTEVPLIEYSFYSD---SLQYPKTVRVVPYKAGAVNPT 274
Qy 246 VTLSLANI--ASKEVKQAPIDAFESTDLIIG-----EVAWLTDTHTTVAAKAFNRVQDQ- 297
Db 275 VKFFVWNTDLSLVNATSIQITAPASMLGDHYLDCVWTWATQER--ISLQWLRRIQNY 332
Qy 298 -QKVAVADTASNATVISDRD---GTDGWLNDLLSMKYIGPIKPS-----DKDAYIID 346
Db 333 VMDICDYDESSGRWNCVLAHQHIEMSTIGM-----VGRFRPSPHFTLDGNSFYKI 383
Qy 347 ISDHSGWAHLVLPVPSGGEPIPLTKGDWEVTSILSDOERQLVYLTQHHSTE--RHLY 404
Db 384 ISNEEYRHCYFQIDKKDCTFITKGTWEVIGIEALTSD--YLYYISNEYKMGMPGGRNLY 441
Qy 405 SVSYSTFA-VTPL---VDDTVAAWNSAFSANSQYIILTYGGPDVPYQELVTTNSTKPLR 460
Db 442 KIQLSDYTKVTCLSCELNPERCQYISVSFSKEAKYQLRCSGGPLPLYTLHSSVNDKGLR 501
Qy 461 TITDNA---KYLEQIK-----DYALPNITYFELPLPSGETLNMVQRLPPGSPDKKYP 511
Db 502 VLEDSALDKMLQNVQMPKSLDFIILNETKFWYQ-----ILPHFDKSKKYP 551
Qy 512 LFTPYGGGAQEVTKRQWALNFKAYVASDSLELYVTWTVNDRGTGFGKGRKFRSAVTRQLG 571
Db 552 LLDVYAGPCSQKADTVFR-LNWAYTLASTENI--IVASFDCGRSGYQGDKIMHAINRRLG 608
Qy 572 LLEAEDQIYAAQAAANIPWIDADHIGWGSFGYLTSLKLEKDSGAFLLGVITAPYSDW 631
Db 609 TFEVEDQIEAARQFSKMGFVNDKRIALWGSYGGYVTSWLGSGGVKFCGIAVAPVSRW 668
Qy 632 RFYDSMYTERYM--KTLSTNEEGYETSAV-RKTDGFKNVEGGFLIQHGTGDDNVHFQNSA 688
Db 669 EYDYSVYTERYMGUPTPEDNLDHYRNSTVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 726
Qy 689 ALVDLLMGDGVSPKLSHQWTFSDHGISYHGGGVFLYKQLARKLYQ 735
Db 727 QISKALVDVGD---FQAMWYTDDEHGIASSTAHOHIYTHMSHFQIK 770
```

Search completed: January 9, 2006, 03:21:31
Job time : 30 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2006, 03:20:31 ; Search time 76 Seconds
(without alignments)
4150.804 Million cell updates/sec

Title: US-10-723-807-2_COPY_17_771
Perfect score: 4033
Sequence: 1 LDVPRKPAPTEGSKRLTF.....EKNRQTQVLHMQHTKDLLE 755

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4033	100.0	771	4	US-10-723-807-2
2	2914	72.3	764	4	Sequence 2, Appli
3	2610.5	64.7	775	5	Sequence 12802, A
4	1269	31.5	790	4	Sequence 35, Appl
5	1162	28.8	793	4	Sequence 3803, Ap
6	1054	26.1	818	4	Sequence 2243, Ap
7	1054	26.1	818	4	Sequence 13304, A
8	1054	26.1	818	4	Sequence 3, Appli
9	1054	26.1	818	4	Sequence 3, Appli
10	1054	26.1	818	4	Sequence 3, Appli
11	1054	26.1	818	4	Sequence 22029, A
12	1054	26.1	818	6	Sequence 3, Appli
13	1054	26.1	818	6	Sequence 3, Appli
14	1054	26.1	818	6	Sequence 3, Appli
15	1054	26.1	818	6	Sequence 3, Appli
16	1054	26.1	818	6	Sequence 3, Appli
17	1020	25.3	504	4	Sequence 863, App
18	987.5	24.5	766	3	Sequence 3, Appli
19	987.5	24.5	766	3	Sequence 1, Appli
20	987.5	24.5	766	4	Sequence 590, App
21	987.5	24.5	766	5	Sequence 18, Appl
22	987.5	24.5	766	5	Sequence 565, App
23	982	24.3	760	5	Sequence 4171, Ap
24	981.5	24.3	766	4	Sequence 6, Appli
25	981.5	24.3	766	4	Sequence 7, Appli
26	981.5	24.3	766	4	Sequence 6, Appli
27	981.5	24.3	766	4	Sequence 922, App

ALIGNMENTS

RESULT 1

US-10-723-807-2
; Sequence 2, Application US/10723807
; Publication No. US20040171104A1
; GENERAL INFORMATION:
; APPLICANT: Alexander Blinkovsky
; APPLICANT: Kimberly Brown
; APPLICANT: Michael W. Rey
; APPLICANT: Alan Klotz
; APPLICANT: Tony Byun
; TITLE OF INVENTION: Polypeptides Having Dipeptidyl
; TITLE OF INVENTION: Aminopeptidase Activity And Nucleic Acids Encoding Same
; FILE REFERENCE: 5254.200-US
; CURRENT APPLICATION NUMBER: US/10/723,807
; CURRENT FILING DATE: 2003-11-25
; PRIOR APPLICATION NUMBER: US/09/079,592
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 08/857,884
; PRIOR FILING DATE: 1997-05-16
; PRIOR APPLICATION NUMBER: 60/062,892
; PRIOR FILING DATE: 1997-10-20
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 771
; TYPE: PRT
; ORGANISM: Aspergillus oryzae
US-10-723-807-2

Query Match 100.0%; Score 4033; DB 4; Length 771;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 755; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LDVPRKPAPTEGSKRLTFNETVVKQAITPTSRVQWLSGAEDGSGYVVAEDGSLTIEN 60
Db 17 LDVPRKPAPTEGSKRLTFNETVVKQAITPTSRVQWLSGAEDGSGYVVAEDGSLTIEN 76
QY 61 IVTNESRTLIPADKIPTGKEAFNYWIHPDLSSVLWASNHTKQYRHSFFADYVQDVESLK 120
Db 77 IVTNESRTLIPADKIPTGKEAFNYWIHPDLSSVLWASNHTKQYRHSFFADYVQDVESLK 136
QY 121 SVPLMPDQEGDIQYQWSPVGNITIAFVRENDLYVWDNGTVTRITDDGGPDMFHVDPDWLY 180
Db 137 SVPLMPDQEGDIQYQWSPVGNITIAFVRENDLYVWDNGTVTRITDDGGPDMFHVDPDWLY 196
QY 181 EEEIIGDRYALWFSFDPGEYLAFLSNFTGVPTTYQYYMNDNOEIPAYPWELKIRYPKVS 240
Db 197 EEEIIGDRYALWFSFDPGEYLAFLSNFTGVPTTYQYYMNDNOEIPAYPWELKIRYPKVS 256
QY 241 QTNPTVTLSLLNIASKEVKQAPIDAFESTDLIIGEVAWLTDTHTTVAAKFNRVQDQKV 300

```
Db 257 QTNPTVTLNLINIAKSEVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDOQKV 316
Qy 301 VAVDTASNKATVISDRDGTGWLNDLLSMKYIGIPKPSDKDQVYIDISDHSGWAHLXLPF 360
Db 317 VAVDTASNKATVISDRDGTGWLNDLLSMKYIGIPKPSDKDQVYIDISDHSGWAHLXLPF 376
Qy 361 VSGGEPILTKGDWEVTSILSIDQERQLVYVYLSLSTQHHSSTERHLYSVSYSTFAVTPVDDT 420
Db 377 VSGGEPILTKGDWEVTSILSIDQERQLVYVYLSLSTQHHSSTERHLYSVSYSTFAVTPVDDT 436
Qy 421 VAAWTSASFNSGYYILTYGGPPVPVQELTYTNTSTKPLRTITDIAKVLQIKDYALPNI 480
Db 437 VAAWTSASFNSGYYILTYGGPPVPVQELTYTNTSTKPLRTITDIAKVLQIKDYALPNI 496
Qy 481 TYFELPLPSGSETLNMORLPPGFPSPDKKYPIILFTPYGPGQAQVTKRWQALNFKAYVASD 540
Db 497 TYFELPLPSGSETLNMORLPPGFPSPDKKYPIILFTPYGPGQAQVTKRWQALNFKAYVASD 556
Qy 541 SELEYVTWTDNRGTGFKGRKFRSAVTRQLGLLEAEQIYAAQQAANI PMIDADHIGWG 600
Db 557 SELEYVTWTDNRGTGFKGRKFRSAVTRQLGLLEAEQIYAAQQAANI PMIDADHIGWG 616
Qy 601 WSPFGYILTSKYLEKDSGAFTLGIVITAPVSDWRFYDSMYTERYMKTLSTNEEGYETSARK 660
Db 617 WSPFGYILTSKYLEKDSGAFTLGIVITAPVSDWRFYDSMYTERYMKTLSTNEEGYETSARK 676
Qy 661 TDGFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSHGSIYHG 720
Db 677 TDGFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSHGSIYHG 736
Qy 721 GGVFLYKQLARKLYQEKNRQTVLMHQTWKDLEE 755
Db 737 GGVFLYKQLARKLYQEKNRQTVLMHQTWKDLEE 771

RESULT 2
US-10-369-493-12802
; Sequence 12802, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: 2003-02-28
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12802
; LENGTH: 764
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(764)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12802

Query Match 72.3%; Score 2914; DB 4; Length 764;
Best Local Similarity 69.13%; Pred. No. 6e-228;
Matches 524; Conservative 103; Mismatches 113; Indels 16; Gaps 4;

Qy 4 PRKHAPTEGSGKRLTFNETVVKQAITPTSRVQSLSGAEDGSGYV-----YAAEDGSLTI 58
Db 17 PWRPREPAAGSKRLTFNETWISAALSPSSISVQWIAENTDGGLCVPGGKWKHQDXEHCH 76
Qy 59 ENIVTNSRTLIPADKIPTGKEAFNWIHPDLSSVLWASNHTKYQRHSFFADYYVQDVES 118
```

```
Db 77 QPIADNRSAEKIPAD-----AYSWISPDLSAVLWATNYTKQYRHSFFADYYIQDVE 129
Qy 119 LKSVLMPDQEGDIQAOQMSPVGNITIAFRENDLYYVMDNGCTVTRITDQGGPDMPHGVDPW 178
Db 130 LETVPLVEDWVGDIQYAEWSPSGDSIAFVRGNLWMTSDGTVAITKDGPGDFMFGHGVDPW 189
Qy 179 IYEEELGDRYALWFPDPGEYLAFLSNFNETGVYTYVQYVMDNQEIAPAYPWEKLIYDPK 238
Db 190 IYEEELGDRFALWFPDSELLAFLTFNETGVPTFTVQYFMDNQEIAPPYPRELDIYPK 249
Qy 239 VSGTNPVTLSLNLINIAKSEVKQAPIDAFESTDLIIIGEVAWLTDTHTTVAAKAFNRVQDOQ 298
Db 250 VSETNPVTKUNLIQLSDNTVSTIPIDVDFPSSELLIYVEVAVWTDTHIELAKAFNRVQDES 309
Qy 299 KVVAVDTASNKATVISDRDGTGWLNDLLSMKYIGIPKPSDKD---AYYIDISDHSGWAH 355
Db 310 KVVIVETASGETKIAHERDGTGWLNDLLSISVVGPLALGSGDASSAYVYVLDLSHSGWTH 369
Qy 356 LYLFPVSGGPIPLTKGDWEVTSILSIDQERQLVYVYLSLSTQHHSSTERHLYSVSYSTFAVTP 415
Db 370 LYLFTSGGPIPLTEGEWEVTSIVSIDQERELFYLSLSTQHHSSTERHLYSVSYRTFTITP 429
Qy 416 LVDDTVAAYWSASFNSGYYILTYGGPDVPVQELTYTNTSTKPLRTITDIAKVLQIKDY 475
Db 430 LVDDTVAAYWSVFSKAGYYILTYAGSPVYQELISVNOQAPLRTILTSNAALIEKLEEY 489
Qy 476 ALPNITYFELPLPSGETLNMORLPPGFPSPDKKYPIILFTPYGPGQAQVTKRWQALNFKA 535
Db 490 ALPNISYFELEIPSGEKLNMORLPVGFPSPDKKYPIILFTPYGPGQAQVSKRWQSLDFNA 549
Qy 536 YVASDSELEVVTWTDNRGTGFKGRKFRSAVTRQLGLLEAEQIYAAQQAANI PMIDADH 595
Db 550 YIASDPELEYVTWTDNRGTGFKGRKFRSLVAKQLGLEAEDQVYAAKQAAKLDWVDSH 609
Qy 596 IGIWGSFGVGLTSKYLEKDSGAFTLGIVITAPVSDWRFYDSMYTERYMKTLSTNEEGYET 655
Db 610 IAIWGSYGGTGLTKVLETSUGAFSLGLLTAPVSDWRLYDSMYTERYMKTLSTNAEGYNT 669
Qy 656 SAVRKTDGFKNVEGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSHG 715
Db 670 TAIHRTDGFKNVEGGFLIQHGTGDDNVHFQNSAALGDTLIGNVTPEKMQVQWFTDSHS 729
Qy 716 ISYHGGVFLYKQLARKLYQEKNRQTVLMHQTWK 751
Db 730 IRYNGGNVFLYRQLAQLRYKEKNRAKKE-QHOMSKR 764

RESULT 3
US-10-926-188A-35
; Sequence 35, Application US/10926188A
; Publication No. US20050158298A1
; GENERAL INFORMATION:
; APPLICANT: Atheris Laboratories
; TITLE OF INVENTION: Novel Fungal Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 28511-501
; CURRENT APPLICATION NUMBER: US/10/926,188A
; CURRENT FILING DATE: 2004-08-25
; PRIOR APPLICATION NUMBER: US 60/498,318
; PRIOR FILING DATE: 2003-08-25
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 35
; LENGTH: 775
; TYPE: PRT
; ORGANISM: Trichophyton rubrum
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (162)
; OTHER INFORMATION: wherein Xaa can be any amino acid
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (612)
```

; OTHER INFORMATION: wherein xaa can be any amino acid
US-10-926-188A-35

```
Query Match      64.7%; Score 2610.5; DB 5; Length 775;
Best Local Similarity 61.6%; Pred. No. 3.2e-203;
Matches 463; Conservative 134; Mismatches 148; Indels 7; Gaps 4;

QY 4 PRKHPTGEGSKRTFTNETVVKQAITPTSRVQWLSGAEQSGYVYAAEDSLTIENVT 63
DB 19 PREPSPTGGGNKLTYYKECVPRATISPRSTSLAMINSEEDGRYISQSDDGALIIQNI 78

QY 64 NESRTLIIPADKIPTCKEAFNWIHPDLSVLWASNHTKQYRHSFPADYYVQDVESLKVP 123
DB 79 NTKTLVAADKVPKG--YYDYFKPDLASVLATNVTYQYRHSYFANYFILDIKKGSLTP 136

QY 124 LMPDEGDIQYQWSPVGNITIAFVRENDLYVNDGTVTRITDDGDPMPHGVDPWYEE 183
DB 137 LAQDAQGDIQYQWSPMNNISIAVVRXNDLYIWNNGKTRKRIENGDPDPNGVDPWYEE 196

QY 184 ILGDRVALWFSDEGEYLAISFNETGVTYTYVQYMDNQEIAPAYPWELKIRYKPVQSTN 243
DB 197 IGDRLPALWFSDEGEYLAISFNETGVTYTYVQYMDNQEIAPAYPRELEIRYKPVSAKN 256

QY 244 PTVTLSLNLASKEVQKAPIDAFESDILIGEVAMLTDTHTTAAKAFNRVDDQKQVAV 303
DB 257 PTVQPHLLNIASSQETTPVTAAPPENDLVIGEVAWLSGHDGSAVAFRNRVQDREKIVSV 316

QY 304 DTASNAKATVISDRDGTGDLNLSKYGIPKPSDKDAYYIDISDHSGWAHLXLPVSG 363
DB 317 KYESKESKIRERDGTGDLNLSKYSYIGNVNGKE---YYVDISDASGWAHIYLPVDG 373

QY 364 GPIPLTKGDWEVTSILSIDORQLVYVLSSTOHHSTERHLYSVSYSTFAVTPLVDDTVAA 423
DB 374 GKEIALTKGEWEVAILKVDTKKLLYFTSTKYHSTRHVSVDTKVMTPLVNDKEAA 433

QY 424 YNSAFSANSYGYIITYGGDPVPYQELYTT--NSTKPLRTITDPAKLEIQDYALPNITY 482
DB 434 YITASFSAGGYIITYSGPNVPYQELYSTKDSKPLKTIITNDALLEKLEKVPKVSF 493

QY 483 FELPLPSGTLNMQRLPGFSPDKYPIPLFTPYGGPGAQEVTKRQALNFKAYVASDSE 542
DB 494 FBIKLPSETNLVKORLPNFPNFKYVPLFTPYGGPGAQEVSOAWNSLDPKSYITSDPE 553

QY 543 LEVVTWVDNRGTGPKGRFSAVTRQLGLAEADQIYAAQQAANIPIWIDAHIGIWS 602
DB 554 LEVVTWVDNRGTGPKGRFSAVTRQLGLAEADQVFAKEVLKRWADKDHIGIWS 613

QY 603 FGGLYTSKVLKDSGAFTLGIVITAPVSDWRPFYDSMYTERYMKTLSTNEEGYETSARVKT 662
DB 614 YGGFLTAKTLETDGVTFTGISTAPVSDFRLYDSMYTERYMKTLVELNADGYSETA 673

QY 663 GPRQVEGGFLIOHGTGDDNHNQNSAALVDLLMGDGVSPKLSQWFTSDHGISYHGG 722
DB 674 GPKNLKGHYLIQHTGDDNHNQNSAALVNTLMNGGVGTADKLTQTQWFTSDHGIYDMS 733

QY 723 VFYKQLARKLYOEKNRQTK--VLMHQTWKOL 753
DB 734 TYQYKQLSKMVDQKORRPESPPMHQSKRVL 765
```

RESULT 4

US-10-369-493-3803
; Sequence 3803, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 3803
; LENGTH: 790
; TYPE: PRT
; ORGANISM: Neurospora crassa
US-10-369-493-3803

Query Match 31.5%; Score 1269; DB 4; Length 790;

Best Local Similarity 38.7%; Pred. No. 5.8e-94;

Matches 296; Conservative 132; Mismatches 271; Indels 66; Gaps 29;

```
QY 7 PHAPTGEKS-KRLTFNETVVKQAITPTSRVQWLSG--AEDG-----SYVYAAD 53
DB 34 POATVTHSGKKVTLQ--VLNNEWRAKSHSISWAGVNGEDGLLEKEGANKDYLVEDV 92

QY 54 GSLTIENIVTNESRTLIPADKIPTCKEAFN--YWIHP-----DLSSVLWASNHTKQYR 104
DB 93 RAQNPSSEVSEAKSKTLI-KDKL---FEFANKTYM--PTVTVPSTRDLKKVLLATDVQNNR 146

QY 105 HSFPADYYVQDVESLSKSVPLMP--DQEGDIQYQWSPVGNITIAFVRENDLYV--WDNGTVT 161
DB 147 HSYAYAVYIIPDVETQQAEPVYDADARLQLASWSPSTDAIVYTRDNNNFKRLSDKLV 206

QY 162 RTTDDGGDPMPHGVDPWYEEIILGDRVALWFSDEGEYLAISFNETGVTYTYVQYTM-- 219
DB 207 QITROGSDADVNGVDPWYEEVEVLASGVATWWSGDNVAVFRTNETGVPPIQYFVSR 266

QY 220 --DNQBIAPA---YPWELKIRYKPVKSQTNPVTLSLNLNLTASKVVKQAPIDA--FESTDLIIG 274
DB 267 PSGEPEKPEENYPEVRQIKYKAGAHNPIVDLKFDVKRGDVFSDISGRFADDDRLIT 326

QY 275 EVAMLTDTHTTVAAKAFNRVQDQKQVAVDTASN-----KATVISDRDGTGDLNLSM 329
DB 327 EVIWM---AGKQVLIKETNRVSDMRVLDVDSRTGKAVRTVVDVNDIDG--GMFEISHKT 381

QY 330 KYIGPIKPSD--KDAYYIDISDHSGWAHL--YLFVPSGGSEPIPLTKGDWEVTSILS-IDOE 385
DB 382 KEI--PADPANGRPDDGYVDVTIHHNGDHLAYFTPLDNPIMPILTSGDYEVDAPSVDLQ 440

QY 386 ROLVYTLSTQHHSTERHLYSVSYSTFAVTPLVDDTVAAAYWASFSANSYGYIITYGGPDV 445
DB 441 RNLVYFVSTKESSIQRHVYQVKLTGEDMTPVTDTSKEGYAISFSTGAGYALVSQGPNI 500

QY 446 PYQELYTT--NSTKPLRTITDPAKLEIQDYALPNITYFELPLPSGETLNMQLPPGF 503
DB 501 PMQKVIPTSPNDPKYEHVVEENKDLAEAKGHELPIINIYGTINV--DGVELNIERRPPHF 559

QY 504 SPDKKYPILFTPYGGPGAQEVTKRQALNFKAYVASDSLEYVVTWVDNRGTGFKGRKFR 563
DB 560 DNKKKYPVLVFOQYSGVPSQTVKKT--AVDFQSFVAAG--LGVICVTVDRGRTGFGIKNR 616

QY 564 SAVTRQLGLLEADQIYAAQQAANIPIWIDAHIGIWSFGGYLTSKVLKDSG--AFTLG 622
DB 617 VLIIRGNLGTWESHDLIAAKHWAQDYIDEDRLAIGWSYGGVYMTLKTLEQDAGQTFKYG 676

QY 623 VITAPVSDWRPFYDSMYTERYMKTLSTNEEGYETSARVKTDPG--KNVEGGFLIOHGTGDDN 681
DB 677 MAVAPVTDWRFYDSIYTERYMERTPQTNPEGYESAATVNTVNTALSQNR--FLMHGVADDN 734

QY 682 VHFQNSAALVDLLMGDGVSPKLSQWFTSDHGISYHGGVPLY 726
DB 735 VHMQNSLTLLDAL--DQRSVENYDVQVFPDSDHGIYFHNANRIVF 777
```

RESULT 5

US-10-369-493-2243
; Sequence 2243, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:

/ APPLICANT: Cao, Yongwei
 / APPLICANT: Hinkle, Gregory J.
 / APPLICANT: Slater, Steven C.
 / APPLICANT: Goldman, Barry S.
 / APPLICANT: Chen, Xianfeng
 / TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 / TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 / FILE REFERENCE: 38-10(52052)B
 / CURRENT APPLICATION NUMBER: US/10/369,493
 / CURRENT FILING DATE: 2003-02-28
 / PRIOR APPLICATION NUMBER: US 60/360,039
 / PRIOR FILING DATE: 2002-02-21
 / NUMBER OF SEQ ID NOS: 47374
 / SEQ ID NO 2243
 / LENGTH: 793
 / TYPE: PRT
 / ORGANISM: Schizosaccharomyces pombe
 / US-10-369-493-2243

Query Match	28.8%	Score 1162	DB 4	Length 793
Best Local Similarity	36.1%	Pred. No. 3e-85		
Matches	272	Conservative 142	Mismatches 294	Indels 46
Gaps	23			
Qy	22	ETVVQKAITTSRSQVWL--SGAEDGSVYAAEAGSLTIENIVNESRTLIP-ADKPIPTG	78	
Db	58	DDIVLQKYKPSYKQVNWIDSQGLKD---TFVLVTKYGDGLINIQQDPYLNKLTLEFSVSDLTVNG	114	
Qy	79	--KEAFNVHTPDLSSVLWASNNHTKQYRHSFADYYVQDVESLKSVPML--PDQEGDIQY	134	
Db	115	IQLDYDSYSISFDKAYVLVSVNKSQRWHSFQAQYLYNTET-KDVNMLGQDNEHWITSL	173	
Qy	135	AQMSPVGNITAFVRENDLYWDN-GTVTRITDDGGPMFHGVPDWIYEEELIGDRYALWF	193	
Db	174	AEMSPTGHQLSFVYVNDLYVRKNDGNVQRLTYDGTVDVFNGLTDMWIEEEVLSPSTIWM	233	
Qy	194	SPDGEYLAYSFNETGYPPTIVQYMDNOEITAPAYP-----WELKIRPKVQSOTNPVTVL	248	
Db	234	SPDSDKIAFLKLNSEIPTY--HYPLTYAELDPLPEFDYKNQMAIKYPKPGNPNPSVSL	291	
Qy	249	SLNIAKSEVKQAIPIDAPES-TDLIIIGEVAMLTDTHTTVAAKAFNRVQDQOKVAVDPTAS	307	
Db	292	FVADENNASNFSLWNEPLAEVPVQVNLVWNT--SSVLVQFTNENSTCITARLLDTEL	349	
Qy	308	NKATVISDRDGTGDLNLLSMKYIGPIKPS-----DKDAYIIDISDHSGWAHLYLFPVS	362	
Db	350	KSIHTVXTECLEEGWYEVQOAKMF-PLNNSLVWENWSDGYF-DIALDDYNHLAFIPFN	407	
Qy	363	GGEPIPLTKGDWEVT-SILSTSQERQLVYVLSQHHSTERHLYSVSYSTFATVPLVDD-T	420	
Db	408	GSSPIYLTSGAWMDVTGDSPIHTDGDGPNVYFATLTKDSTERHLYVYVSLDTLBIYGITDNGE	467	
Qy	421	VAAWKSASFNSANGYIILTYGGPDVPYQELTYTNTSKPLRITITONAKVLEQIKQVYALNNI	480	
Db	468	DEGYSTSFSPFGDFYVLVNHGPDVPWQELASTKDKQCYCLESUETNSRUKQQLSSITLPSV	527	
Qy	481	TYFELPLPSGDTLNVQRLPPGSPDKKYPILFTPYGGPGAQEVTKRQWALNFKAYVASD	540	
Db	528	EYKLTLP-NDTTFNFMERRPNFDVNKKYPVLFFAYGGGSGQOAKLFR-VDFQAYLASH	585	
Qy	541	SELEYVTWVDNRGTGFKGRKFRSAVTRQLGLEABDQIYAAQQAANTIPWIDAHIGITWG	600	
Db	586	PDFEFIVVTLDRGTGFGNAGAFRYSVSRHLGEWESYDQOQAGKFWADLPFVDENHVGITWG	645	
Qy	601	WSFGYLTSKVLKSDSGAFTLGVTAPYSDRRFPYDSMYTERVMKTLSTNEGEYETSAVRK	660	
Db	646	WSYGGYLLTKLTLETD-VFSYGMAYVATDRLYDSVYTERMDLPQNKGEYKNSQIHD	704	
Qy	661	TGDFKNVGGFLIOHGTGDDNVHFONSAAALVDLLMGDGVSPBKLSH---QWFTDSDHGIS	717	
Db	705	YEKFKQLK-RFFVAHGTGDDNVHFGHSHML-----DGLNLANCNYDMAVFPDPSAHSIS	758	
Qy	718	YHGGGVFLYKQLARKLYQOEKNR---QTQVLMHQW	748	

DB 759 YHNASLSIYHRLSEWIGDALGRIDPSTGVROHRW 792

RESULT 6

US-10-369-493-13304

; Sequence 13304, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 13304

; LENGTH: 634

; TYPE: PRT

; ORGANISM: Aspergillus nidulans

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)..(634)

; OTHER INFORMATION: unsure at all Xaa locations

US-10-369-493-13304

Query Match	27.7%	Score 1118;	DB 4;	Length 634;
Best Local Similarity	38.1%;	Pred. No. 8.4e-82;		
Matches	246;	Conservative 105;	Mismatches 237;	Indels 58; Gaps 17
QY	129	EGDIQYQAWGSPVGNITAFVRENDLYV--WDMGTVTRITDDGGPDMFGVDPWDIYEEELG	186	
DB	2	DGRVQLALWLPNSDAIAFVRDNNLYLRKLSSERVVPIITKDGGEQLFGVDPWDIYEEVFS	61	
QY	187	DRYALWSPDGEYLAFLSNETGVTPTTVQYYMDN-----OETAPAYPWLKLYRYPKVS	240	
DB	62	GNSVTWMSDGSQIAFIRTNESAAPVPEFVQYFLSRSPGKKPQGLENYPEVREIKYIPKAG	121	
QY	241	QTNFTVTLISLINIASKVKQAPI--DAFESTDLIIIEVAWLTDHTTAAKAFNRVQDOOK	299	
DB	122	APNFVNLFQVDEQGEVFSVDTFDDDDRLIIIEVWA--AKGKVLVRTTNRESDLK	179	
QY	300	VVAVDITASNKATVYISDRDGT-----GWLNDLLSMKYIGIPKPSDKDA-----YYIDISHS	351	
DB	180	VFLVDTESRESKLIRIQDIAELDGGWVEPTQSVRFI----PADPDKGRFPDGYLDTVVHE	235	
QY	352	GWAHL--YLFPVSGGCEPIPLTKGDMEVTSI--LSIDQEROLVYVLSLTOHSTHERHLYSVSYS	409	
DB	236	GYDHLAYFTPLNDNEPIMLTSGEWEVDVDAPTAVDLTRGLVVFIAATKEAPTEHLRYRDL	295	
QY	410	TFAVTPLVDDTVAAYWASFSANSYXILTYGGDPVPYQELVYT--NSTKPLRTITDNAK	467	
DB	296	GSDLTPLDTSQPGYVSVSPSDGAGYALLSYQGPSIPWQSIIISTEGEKTTTLRIIEDNTD	355	
QY	468	VLEQIKDYALPNITYFELPLPSGETLNVQRLPGCFSPDKVKYPILFTPYGGPGGAQEVTKR	527	
DB	356	LSKLVAQYALPTENYQNITI--DGFTLQVVERRRPHFNENPARKYPLVFLHYGGPGSTVDNR	414	
QY	528	WOALNFKAIVASDSLELYVTWVDNRGTGFKGRFRSAVTRQLGLLEAEDQIYAAQQAAN	587	
DB	415	FN--VDFQSYVA--ASLGIYIVTVDRGGTGFGIGRAARCIIRGNIGHYEAIQDIATAKNNAQ	471	
QY	588	IPWIDADHIGWGSFGYLTSKLEKDSG--AFTLGVITAPVSDKRRFYD-----	635	
DB	472	KPYVDESMAIWGWSYGFMTLKLLEQDAGETFOYGMVAVPDWMRFYTGKALKPTHFLLG	531	
QY	636	-----SMYTERYMKTLSTNEBGETSAVRKTDGPKNVBGGFLIQHGTGDDNVH	683	

Query Match	27.7%	Score 1118;	DB 4;	Length 634;
Best Local Similarity	38.1%;	Pred. No. 8.4e-82;		
Matches	246;	Conservative 105;	Mismatches 237;	Indels 58; Gaps 17
QY	129	EGDIQYQAWGSPVGNITAFVRENDLYV--WDMGTVTRITDDGGPDMFGVDPWDIYEEIEILG	186	
DB	2	DGRVQLALWLPNSDAIAFVRDNNLYLRKLSSERVVPIITKDGGEQLFGVDPWDVVEEVEFS	61	
QY	187	DRYALWSPDGEYLAFLSNETGVTPTTVQYYMDN-----OETAPAYPWELKLYRYPKVS	240	
DB	62	GNSVTWMSDGSQIAFIRTINESAVPEFVQVFLSRPQKPKQGLENYPEVREIKYIPKAG	121	
QY	241	QTNFTVTLISLINIASKVKQAPI-DAPFESTDLIIIEVAWLTDHTTAAKAFNRVQDOOK	299	
DB	122	APNFVNLFQVDEQGEVFSVDTFDDDDRLIIIEVWA--AKGKVLVRTTNRESIDLK	179	
QY	300	VVAVDITASNKATVYISDRDGT-----GWLNDLLSMKYIGIPIKPSDKDA-----YYIDISHS	351	
DB	180	VFLVDTESRESKLIRIQDIAELDGGWVEPTQSVRFI----PADPDKGRFPFGDYLTVVHE	235	
QY	352	GWAHL-YLFPVSGGCEPIPLTKGDMEVTSI-LSIDQEROLVYVLSLTOHSTHERHLYSVSYS	409	
DB	236	GYDHLVFTPLDNPPEPIMLTSGEWEVDVDAPTAVDLTRGLVVFIAATKEAPTEHLRYRVL	295	
QY	410	TFAVTPLVDDTVAAYWASFSANSYXILTYGGDPVFPQELVYT--NSTKPLRTITDNAK	467	
DB	296	GSDLTPLTDTSQPGVYSVPSDGGAYALLSYQGPSIPWQSIIISTEGEKTTTLRIIEDNTD	355	
QY	468	VLEQIKDYALPNITYFELPLPSGETLNVQRLPGCFSPDKVKYPILFTPYGGPGGAQEVTKR	527	
DB	356	LSKLVAQYALPTENYQNITI-DGFTLQVVERRRPHFNENPARKYPLVFLHYLGGPGSTVDNR	414	
QY	528	WQALNFKAIVASDSLELYVTWVDNRGTGFKGRFRSAVTRQLGLLEAEDQIYAAQQAAN	587	
DB	415	FN-VDFQSYVA--ASLGIYIVTVDRGGTGFGIGRAARCIIRGNIGHYEAIQDIATAKNNAQ	471	
QY	588	IPWIDADHIGWGSFGYLTSKLEKDSG-AFTLGVITAPVSDKRRFYD-----	635	
DB	472	KPYVDESMAIWGWSYGFMTLKLQDAGETFOYGMVAVPDWRFTGKALKPTHFLLG	531	
QY	636	-----SMYTERYMKTLSTNEBGETSAVRKTDGPKNVBGGFLIQHGTGDDNVH	683	


```
Db 532 PCIVNKPRAQXISVYTERYHTPOHNPCTGYDNTSISDMAALHN-NVRFVLHIGASDDNVH 590
Qy 684 PQNSAALVDLLMGDGVSPKLSHQWFTSDHGIGSYHGSGVFLYKOL 729
Db 591 IQNTLTLDKUL-DLASQNVYDVHFPDSDHSIFFHNAHTWVIERM 634

RESULT 7
US-10-401-437-3
; Sequence 3, Application US/10401437
; Publication No. US20030158070A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/10/401,437
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PR1
; ORGANISM: Bacillus subtilis
US-10-401-437-3

Query Match 26.1%; Score 1054; DB 4; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
Db 48 IPHNSNTPDYQBPNSNYTNDGKLVKVSFS-VVRNNTFPQPKYHBLQWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNESRTLLIPADKIPGKEAFNYWIH-----PDLSSVLW 95
Db 107 TPMND-SYVVKSVYDDSYNSVLLEGKT-----FIHQNLTVESITASPDLLKLLI 156
Qy 96 ASNHTKQYRHSFPADYVVDVESLKSVPMLPQDEGDIQVAQSPVGNITAFVRENDLYW 155
Db 157 RINSQVNRHSITFGSYFYVD---KSSSFEEIGNEVALAIWSPNSNDIAYQDNNIYIY 212
Qy 156 ---DNGTVTRITDDGPDMPHGVDPWIYEEILGDRYALWFPDGEYLAFLSFNETGVPT 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAWSPDGYLAFLKIDSEVGE 272
Qy 213 YTVQYMDNQEIAPAPWELKIRYKVSQTNPTVTLSSLNIAKESVQKQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIYPMRSIKYKSGTFNPHAEIWMYKMDGTSFHPRIKSGKDGSL 330
Qy 272 IIGEVAVLTDTHTTVAAKAFNRVQDQKQVAVDTSNKAATVISDRDGTGWLNDLLSMKY 331
Db 331 LITEVTWNGN--VLVKTDRSSDILTFLDITIAKTSNVVRNESSNGGWEITHNTLF 388
Qy 332 IGIKPSDKDAY-YIDISDHSGWAHLYLPVSGGEPI-PLTKGDWEVTS-ILSIDERQL 388
Db 389 IPANETFDPHNGYVDILPIGGYNHLYAFENSNSHYKTLTEGKEVWNGPLAFDSMENR 448
Qy 389 VYVLSQTHSTERHLYSVSY-STFAVTPLVDDTVAAVWSASFSANGSYVILTYGGPDVPY 447
Db 449 LPISTRKSTSRHRYIIDLRSNIEIETDTSDDGYDVYFSGRRFGLLTYKGPVY 508
Qy 448 QELYTNTST-----KPLRTITONAKVLEQIKDYALPNITYFELPLPSGE----491
Db 509 QKIVDHSKAEKCDKGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFRELNLKDFGKD 568
Qy 492 -TLNWMORLPPGSP--DKKYPILFTPYGPGGAQEVTKRQWALNFKAYVASDSELEYVTV 548
Db 569 ILVNSVEILLPNDPDETLSDHPYVFFPAYGPNQSVVKTFF-SVGFNEVVA--SOLNAIVV 625
```

```
Qy 549 TVDNRGTGFKGRKFRSAVTRQIGLLEARDIYAAQAAANIPIWDADHIGIWCWSEGGYLT 608
Db 626 VVDGRGTGFKQDPRSLVRDLRGDYEARDOIQAASLYGSLTFVDPQKISLFQWSEGGYLT 685
Qy 609 SKVLEKDSGA-FTLGVITAPVSDNRFDYSMTYTERYMKLTSTNEEGYETSAVRKTDGFKNV 667
Db 686 LKLTLEKQGRHFKYGMVAPVTDNRFDYSVYTERYMKHTPOENFDGYVSSVHNVTALAQA 745
Qy 668 EGGFLIQHTGDDNVHFNQNSAALVDLLMGDGVSPKLSHQWFTSDHGIGSYHGSGVFLYK 727
Db 746 -NRFLMGTGDDNVHFNQNSLKFLLDLLLNGVENVYDVHV--FPDSDHSIRYHNANVVPD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 8
US-10-402-312-3
; Sequence 3, Application US/10402312
; Publication No. US20030166248A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/10/402,312
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PR1
; ORGANISM: Bacillus subtilis
US-10-402-312-3

Query Match 26.1%; Score 1054; DB 4; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
Db 48 IPHNSNTPDYQBPNSNYTNDGKLVKVSFS-VVRNNTFPQPKYHBLQWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNESRTLLIPADKIPGKEAFNYWIH-----PDLSSVLW 95
Db 107 TPMND-SYVVKSVYDDSYNSVLLEGKT-----FIHQNLTVESITASPDLLKLLI 156
Qy 96 ASNHTKQYRHSFPADYVVDVESLKSVPMLPQDEGDIQVAQSPVGNITAFVRENDLYW 155
Db 157 RINSQVNRHSITFGSYFYVD---KSSSFEEIGNEVALAIWSPNSNDIAYQDNNIYIY 212
Qy 156 ---DNGTVTRITDDGPDMPHGVDPWIYEEILGDRYALWFPDGEYLAFLSFNETGVPT 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAWSPDGYLAFLKIDSEVGE 272
Qy 213 YTVQYMDNQEIAPAPWELKIRYKVSQTNPTVTLSSLNIAKESVQKQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIYPMRSIKYKSGTFNPHAEIWMYKMDGTSFHPRIKSGKDGSL 330
Qy 272 IIGEVAVLTDTHTTVAAKAFNRVQDQKQVAVDTSNKAATVISDRDGTGWLNDLLSMKY 331
Db 331 LITEVTWNGN--VLVKTDRSSDILTFLDITIAKTSNVVRNESSNGGWEITHNTLF 388
Qy 332 IGIKPSDKDAY-YIDISDHSGWAHLYLPVSGGEPI-PLTKGDWEVTS-ILSIDERQL 388
Db 389 IPANETFDPHNGYVDILPIGGYNHLYAFENSNSHYKTLTEGKEVWNGPLAFDSMENR 448
Qy 389 VYVLSQTHSTERHLYSVSY-STFAVTPLVDDTVAAVWSASFSANGSYVILTYGGPDVPY 447
```

```
Db 449 LYFISTRKSTERHVVYIDLRSPNEIIEVTDSEDGVYDVSFSSGRFGLITYKGPVPY 508
Qy 448 QELYTTNST-----KPLRTITDIAKVLQIKDYALPNITTYFELPLPSGE---- 491
Db 509 QKIVDFHSRKAECDCGNVLGKSLYHLEKNEVLTKILEDYAVPRKSPRELNLGKDFGKD 568
Qy 492 -TLNMORLPPGFS--DKKYPILFTPYGGGAQEVTRKWAQALNFKAYVASDSLEYVTW 548
Db 569 ILVNSYELPNDPDETLSDHYPVFFAYGGPNSQOVVKT--SVGFNEVVA--SQNALIV 625
Qy 549 TVDNRGTFGKGRKFRSAVTRQLGLEAEADQIYAAQQAANIPIWDADHIGWGSFGGYLT 608
Db 626 VVDRGRTGFGQDFRSLVRDLRGDYEARQDLSAASLYGSLTFVDPQKISLFGWSYGGYLT 685
Qy 728 QL 729
Db 803 KL 804

RESULT 9
US-10-402-067-3
; Sequence 3, Application US/10402067
; Publication No. US20030175892A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; CURRENT APPLICATION NUMBER: US/10/402,067
; PRIOR FILING DATE: 2003-03-26
; PRIOR FILING DATE: 1998-07-14
; PRIOR FILING DATE: 1998-07-15
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-402-067-3

Query Match 26.1%; Score 1054; DB 4; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGRKLTFTNETVTVVQKAITPSTRSVQWLS-----GAEDGSYV 48
Db 48 IPHHSNTPDYQEPNSNYTNDGKLKVSFS--VVRNNTFPQKYHELOWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNESRTILPADKIPGKEAFNYIHI-----FHHNGQLTVESITASPDLKRLLI 95
Db 107 TFMND-SYVVKSVYDDSYNSVLLGKT-----FHHNGQLTVESITASPDLKRLLI 156
Qy 96 ASNHTKQYRHGFFADYVYQDVESLKVPLMPDQEGDIOYAQWSPVGNITAFVRENDLYW 155
Db 157 RTNSVQNRHSTFGSYFVVD---KSSSFEIENIGNEVALAIWSPNSNDIAYVQDNNIYI 212
Qy 156 ---DNGTVTRITDDCGPMFCHQVPMWIEEELGDRYALWSPDCEYLAYLSFNETGYPT 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPMWYEEEFDDKAAMWSPTGDYLAFLKIDSEVGE 272
Qy 213 YTVQYMDNOETAPAYPELKIYKVSQTNPTVTLSLINTASKEVKQAPIDAFEST-DL 271
Db 273 FIIPYVQDEK--DIYPMERISIKYKSGTTPNPAELWVYSMDKDTSFHPRISGNKKGSL 330
```

```
Qy 272 IIGEVAULTDTHITVAAKAFNRVQDQKVVAVDTASNKATVISDRDGTGDLNLLSMKY 331
Db 331 LITEVTVWNGN--VLVKTTRDSRSDILTVFLDITIAKTSNVVRNNESSNGHWEITHNLT 388
Qy 332 IGPIKPSDKDAY-YIDISDHSWAHLVLPFVSGGEPI-PLTKGDWEVTS-ILSIDQERQL 388
Db 389 IPANETFDPRHNGYVDILPIGGYNHLAYFENSNSHYKTLTEGKWEVVGPLAFDSMENR 448
Qy 389 VYLSSTQHHSTERHLYSVSY-STFAVTPLVDDTVAAVWASAFSANSYIILTYGPDVPY 447
Db 449 LYFISTRKSTERHVVYIDLRSPNEIIEVTDSEDGVYDVSFSSGRFGLITYKGPVPY 508
Qy 448 QELYTTNST-----KPLRTITDIAKVLQIKDYALPNITTYFELPLPSGE---- 491
Db 509 QKIVDFHSRKAECDCGNVLGKSLYHLEKNEVLTKILEDYAVPRKSPRELNLGKDFGKD 568
Qy 492 -TLNMORLPPGFS--DKKYPILFTPYGGGAQEVTRKWAQALNFKAYVASDSLEYVTW 548
Db 569 ILVNSYELPNDPDETLSDHYPVFFAYGGPNSQOVVKT--SVGFNEVVA--SQNALIV 625
Qy 549 TVDNRGTFGKGRKFRSAVTRQLGLEAEADQIYAAQQAANIPIWDADHIGWGSFGGYLT 608
Db 626 VVDRGRTGFGQDFRSLVRDLRGDYEARQDLSAASLYGSLTFVDPQKISLFGWSYGGYLT 685
Qy 609 SKVLEKDSGA-FTLGVITAPVSDWRFYDSMYTERYMKTLSTNERGYETSAVRKTDGPKV 667
Db 686 LKTLKDGGRHFKYGMSPVPTDWRFYDSVYTERYMTPTQENFQYVYESSVHNVTALQA 745
Qy 668 EGGFLIOHGTGDDNVHFNQNSAALVDLLMGDGVSPKLSHQWFTSDHGISYHGGGVFLYK 727
Db 746 -NRELLMHGTGDDNVHFNQNSLKFLDLLDLNGVENYDVHV--FPDSHDSIRYHNANVIFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 10
US-10-401-436-3
; Sequence 3, Application US/10401436
; Publication No. US20030180932A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/10/401,436
; CURRENT FILING DATE: 2003-03-26
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-10-401-436-3

Query Match 26.1%; Score 1054; DB 4; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGRKLTFTNETVTVVQKAITPSTRSVQWLS-----GAEDGSYV 48
Db 48 IPHHSNTPDYQEPNSNYTNDGKLKVSFS--VVRNNTFPQKYHELOWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVNESRTILPADKIPGKEAFNYIHI-----FHHNGQLTVESITASPDLKRLLI 95
Db 107 TFMND-SYVVKSVYDDSYNSVLLGKT-----FHHNGQLTVESITASPDLKRLLI 156
Qy 96 ASNHTKQYRHGFFADYVYQDVESLKVPLMPDQEGDIOYAQWSPVGNITAFVRENDLYW 155
Db 96 ASNHTKQYRHGFFADYVYQDVESLKVPLMPDQEGDIOYAQWSPVGNITAFVRENDLYW 155
```

Db 157 RTNSVQNRHSTFGSVFYVD-----KSSSFEBIGNEVALAIWSPNSNDIAYVQDNIIYI 212
Qy 156 ---DNGTTRITDGGPDGMFHVGPWMIYEEELGDRYALWSPDCEYLAISFNFTGVT 212
Db 213 SAISKKTTRAVTNDGSSFLFNKPKDWMVYEEEFEDDKAAWWSPTGDIYLAFLKIDSEVGE 272
Qy 213 YTVQYMDNQETAPAYPMELKIRYPKVSTNPTVTLISLNIAKSEVKQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIYPEMRSIKYPKSGTPNPHAEIWWYSMDKGTGTFHPRISGNKKGSL 330
Qy 272 IIGEVAMLTDDHTTVAAKAFNRVQDQKVAVDTASNKATVLSDRDGTGWLNDLLSMKY 331
Db 331 LITEVTWVGNG--VLVKTDRSSDILTAVFLDITIAKTSNVVRNENSSNGMWETHTLTF 388
Qy 332 IGPIKPSDKDAY-YTIDSHSGWAHLYLFPVSGGEPI-PLTKGDWEVTS-ILSIDOERQL 388
Db 389 IPANETFRPHNGYVDILPIGGYNHAYFENSNSHYKLTTEGKWEVNGPLAFDSMENR 448
Qy 389 YVYLSTQHHSTERHLYSVSY-STFAVTPLVDDTVAAYWSASFSANSYVILTYGGPDVY 447
Db 449 LFISTRKSTSRHVYIDLRSPNEIIEVTDTSDEGVYDVVSFSSGRRFGLLYTKGPKVY 508
Qy 448 QELYTNTST-----KPLRTITDNAKVLBOIKDYALPNITYFELPLPSGE---- 491
Db 509 QKIVDFHRSKAEKCDKGNVLSKLYHLEKNEVLTKILEYAVPRKSFRELNLGKDFGKD 568
Qy 492 -TLNVMORLPFGFSP--DKKYPILPTPYGCGAQEVTKRQWALNFKAYVASDSELEYVTW 548
Db 569 ILVNSYEILPNDFDETLSDHYVFPFAYGPGNSQQVVKTF-SVGFNEVVA--SOLNAIVV 625
Qy 549 TVDNRGTGPKGRKFRSAVTRQGLLEAEQIIVAAQQAANIPIWDADHIGWMSFGGYLT 608
Db 626 VVDGRGTGFGQDFRSLVRDLGDEYARDQISAAISLYGSLTFVDPQKISLFGWSYGGYLT 685
Qy 609 SKVLEKDSGA-FTLGVITAPVSDWRFYDSMYTERYMKTLSTNEEGYETSARVKTDFGNV 667
Db 686 LKTEKOGGRHFKYGMVAPVTDWRFYDSVYTERYMHHTPOENFDGYVSSVHNVTALAQA 745
Qy 668 EGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSDHGISHYGGVFLYK 727
Db 746 -NRFLMHGTGDDNVHFQNSKFLDLDDLNGVENYDVHV--FPDSDHSIRYHNAVIVFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 11

US-10-369-493-22029
; Sequence 22029, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 22029
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-22029

Query Match 26.1%; Score 1054; DB 4; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;

Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;
Qy 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRSVQMLS-----GAEDGSYV 48
Db 48 IPHNSNTDYQEPNSNYTNDGKLKVSFS-VVRNNTFQPKYHELMQISDNKIESNDLGLYV 106
Qy 49 YAAEDGSJTNIENVTNESRTLIPADKIPITGKEAFNYWH-----PDLSSVLW 95
Db 107 TFMND-SYVVKSVYDDSYNSVLEGT-----FIHNGQLTVSTASPDKRLLI 156
Qy 96 ASNHTQVHRHSPFADYVQDVESLAKSVPLMPDQEGDIOVAQWSPVGNTIAFVRENDLYYW 155
Db 157 RTNSVQNRHSTFGSVFYVD-----KSSSFEBIGNEVALAIWSPNSNDIAYVQDNIIYI 212
Qy 156 ---DNGTTRITDGGPDGMFHVGPWMIYEEELGDRYALWSPDCEYLAISFNFTGVT 212
Db 213 SAISKKTTRAVTNDGSSFLFNKPKDWMVYEEEFEDDKAAWWSPTGDIYLAFLKIDSEVGE 272
Qy 213 YTVQYMDNQETAPAYPMELKIRYPKVSTNPTVTLISLNIAKSEVKQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIYPEMRSIKYPKSGTPNPHAEIWWYSMDKGTGTFHPRISGNKKGSL 330
Qy 272 IIGEVAMLTDDHTTVAAKAFNRVQDQKVAVDTASNKATVLSDRDGTGWLNDLLSMKY 331
Db 331 LITEVTWVGNG--VLVKTDRSSDILTAVFLDITIAKTSNVVRNENSSNGMWETHTLTF 388
Qy 332 IGPIKPSDKDAY-YTIDSHSGWAHLYLFPVSGGEPI-PLTKGDWEVTS-ILSIDOERQL 388
Db 389 IPANETFRPHNGYVDILPIGGYNHAYFENSNSHYKLTTEGKWEVNGPLAFDSMENR 448
Qy 389 YVYLSTQHHSTERHLYSVSY-STFAVTPLVDDTVAAYWSASFSANSYVILTYGGPDVY 447
Db 449 LFISTRKSTSRHVYIDLRSPNEIIEVTDTSDEGVYDVVSFSSGRRFGLLYTKGPKVY 508
Qy 448 QELYTNTST-----KPLRTITDNAKVLBOIKDYALPNITYFELPLPSGE---- 491
Db 509 QKIVDFHRSKAEKCDKGNVLSKLYHLEKNEVLTKILEYAVPRKSFRELNLGKDFGKD 568
Qy 492 -TLNVMORLPFGFSP--DKKYPILPTPYGCGAQEVTKRQWALNFKAYVASDSELEYVTW 548
Db 569 ILVNSYEILPNDFDETLSDHYVFPFAYGPGNSQQVVKTF-SVGFNEVVA--SOLNAIVV 625
Qy 549 TVDNRGTGPKGRKFRSAVTRQGLLEAEQIIVAAQQAANIPIWDADHIGWMSFGGYLT 608
Db 626 VVDGRGTGFGQDFRSLVRDLGDEYARDQISAAISLYGSLTFVDPQKISLFGWSYGGYLT 685
Qy 609 SKVLEKDSGA-FTLGVITAPVSDWRFYDSMYTERYMKTLSTNEEGYETSARVKTDFGNV 667
Db 686 LKTEKOGGRHFKYGMVAPVTDWRFYDSVYTERYMHHTPOENFDGYVSSVHNVTALAQA 745
Qy 668 EGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHQWFTDSDHGISHYGGVFLYK 727
Db 746 -NRFLMHGTGDDNVHFQNSKFLDLDDLNGVENYDVHV--FPDSDHSIRYHNAVIVFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 12

US-11-014-051-3
; Sequence 3, Application US/11014051
; Publication No. US20050095683A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/11/014,051
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15


```
Db 746 -NRFLMHGTGDDNVHFQNSLKFLDLLDNGVENVYDVHV--FPDSHHSIRYHNANVIVFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 14
US-11-014-339-3
; Sequence 3, Application US/11014339
; Publication No. US20050142638A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/11/014,339
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US 09/462,845
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-11-014-339-3

Query Match 26.1%; Score 1054; DB 6; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
Db 48 IPHHSNTPDYQPNNGNYNDGKLVKVSFS-VVRNNTFPQKYHELQWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVTNESRTLIPADKIPTGKEAFNYIH-----FIHNGQLTVESITASPDLLKRLLI 156
Db 107 TFMND-SYVVKSVYDSDSYNSVLEGGT-----FIHNGQLTVESITASPDLLKRLLI 156
Qy 96 ASNHTKQVHSPFADYVQDVESLKSVPMPQEGDIQYAAQSPVGNITAFVRENDLYVW 155
Db 157 RTNSVQNWHRSTFGSYFYVD---KSSSFEEIGNEVALAINSPNSNDIAYVDNNIYIY 212
Qy 156 ---DNGTVTRITDDGPDMPFHGVPDWIYEEELGDRYALWFGPDGEGYLAYLFSNFTGVP 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAWNSPTGYLAFKLIDSEVGE 272
Qy 213 YTVQYMDNQEIAPYAPWELKIRYKVSQTNPTVTLSLLNIASKEVKQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIYPEMRSIKYPKSGTPNPHAEWLWVYSMDKDGTSFHPRISGNKKDGL 330
Qy 272 IIGEVAWLTDTHTTAAKAFNRVQDQKVAVDTSNKAATVISDRDGTGWLNDLLSMKY 331
Db 331 LITEVTWVNGN--VLVKTTRDSSDILTFLDITIAKTSNVVRNNESSNGGWEITHNTL 388
Qy 332 IGPIKPSDKDAY-YIDISDHSQWHAHLYLFPVSGGEPI-PLTKGDWEVTS-ILSIDQERQL 388
Db 389 IPANETFDPHNGYVDILPIGGYNHLAYFENSNSHYKTLTEKWEVNGPLAFDSMENR 448
Qy 389 VYVLSTQHSHTERHLYSVS-SFAVTPLVDDTVAAAYWSASFSANSYILTYGGPDVPY 447
Db 449 LPISTRKSTERHYVYIDLRSENEIIEVTDISEDGVYDVSFSGRRFGLLYTKGPKVY 508
Qy 448 QELYTTNST-----KPLRTITDNKAVLEQIKOYALPNIITYFELPLPSGE---- 491
Db 509 QKIVDPHSRKAECCKDGNVLGKSLYHLEKNEVLTKILEDYAVPRKSFREINLGKDFGKD 568
Qy 492 -TLNVNQRLPPGESP--DKKYPILFTPYGGPGQAQVTKRQALNFKAYVASDSELYVTW 548
```

```
Db 569 ILVNSYEILPNDPDETLSDHYFVFFFAYGGNSQOVVKTF-SVGNEVVA--SQLNAIVV 625
Qy 549 TVDNRGTCFKGRKERSAVTROGLLEABEDQIYAAQQAANI PMIDADHIGIWMGSGGGYLT 608
Db 626 VVDGRGTGFKGQDFRSLVDRDLGDYEARDQISASLYGSLTFDVPQKISLFGWSYGGYLT 685
Qy 609 SKVLEKDSGA-FTLGVITAPVSDMRFRYDSMYTERYMKTLSTNEBGEYTSAVRKTGDFKNV 667
Db 686 LKTLKOGGRHFKYGMVAPVTDWRFYDSVYTERYMHTPQENFDGYVSVSSVHNVTALQA 745
Qy 668 EGGFLIQHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHOMFTDSDHGISVHGGGVFLYK 727
Db 746 -NRFLMHGTGDDNVHFQNSLKFLDLLDNGVENVYDVHV--FPDSHHSIRYHNANVIVFD 802
Qy 728 QL 729
Db 803 KL 804

RESULT 15
US-11-014-364-3
; Sequence 3, Application US/11014364
; Publication No. US20050214616A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC382-US
; CURRENT APPLICATION NUMBER: US/11/014,364
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US 09/462,845
; PRIOR FILING DATE: 2000-01-13
; PRIOR APPLICATION NUMBER: PCT/US98/14647
; PRIOR FILING DATE: 1998-07-14
; PRIOR APPLICATION NUMBER: EP 97305237.7
; PRIOR FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 818
; TYPE: PRT
; ORGANISM: Bacillus subtilis
US-11-014-364-3

Query Match 26.1%; Score 1054; DB 6; Length 818;
Best Local Similarity 33.1%; Pred. No. 2e-76;
Matches 259; Conservative 132; Mismatches 311; Indels 80; Gaps 23;

Qy 3 VPRKPHAP-----TGEGSKRLTFNETVVKQAITPTSRVQWLS-----GAEDGSYV 48
Db 48 IPHHSNTPDYQPNNGNYNDGKLVKVSFS-VVRNNTFPQKYHELQWISDNKIESNDLGLYV 106
Qy 49 YAAEDGSLTIENIVTNESRTLIPADKIPTGKEAFNYIH-----FIHNGQLTVESITASPDLLKRLLI 156
Db 107 TFMND-SYVVKSVYDSDSYNSVLEGGT-----FIHNGQLTVESITASPDLLKRLLI 156
Qy 96 ASNHTKQVHSPFADYVQDVESLKSVPMPQEGDIQYAAQSPVGNITAFVRENDLYVW 155
Db 157 RTNSVQNWHRSTFGSYFYVD---KSSSFEEIGNEVALAINSPNSNDIAYVDNNIYIY 212
Qy 156 ---DNGTVTRITDDGPDMPFHGVPDWIYEEELGDRYALWFGPDGEGYLAYLFSNFTGVP 212
Db 213 SAISKKTIRAVTNDGSSFLFNGKPDWVYEEVEFDDKAAWNSPTGYLAFKLIDSEVGE 272
Qy 213 YTVQYMDNQEIAPYAPWELKIRYKVSQTNPTVTLSLLNIASKEVKQAPIDAFEST-DL 271
Db 273 FIPIYVQDEK--DIYPEMRSIKYPKSGTPNPHAEWLWVYSMDKDGTSFHPRISGNKKDGL 330
Qy 272 IIGEVAWLTDTHTTAAKAFNRVQDQKVAVDTSNKAATVISDRDGTGWLNDLLSMKY 331
Db 331 LITEVTWVNGN--VLVKTTRDSSDILTFLDITIAKTSNVVRNNESSNGGWEITHNTL 388
Qy 332 IGPIKPSDKDAY-YIDISDHSQWHAHLYLFPVSGGEPI-PLTKGDWEVTS-ILSIDQERQL 388
```


GenCore version 5.1.6

Copyright (c) 1993 - 2006 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 9, 2006, 03:21:01 ; Search time 19 Seconds
(without alignments)
337.510 Million cell updates/sec

Title: US-10-723-807-2_COPY_17_771

Perfect score: 4033

Sequence: 1 LDVPRKPAPTGGSKRLTF.....EKNRQTQVLMHQWTKDLEE 755

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 61141 seqs, 8493638 residues

Total number of hits satisfying chosen parameters: 61141

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New:*

1: /cgn2_6/prodata/1/pubpaa/US08_NEW_PUB.pep.*

2: /cgn2_6/prodata/1/pubpaa/US06_NEW_PUB.pep.*

3: /cgn2_6/prodata/1/pubpaa/US07_NEW_PUB.pep.*

4: /cgn2_6/prodata/1/pubpaa/PCT_NEW_PUB.pep.*

5: /cgn2_6/prodata/1/pubpaa/US09_NEW_PUB.pep.*

6: /cgn2_6/prodata/1/pubpaa/US10_NEW_PUB.pep.*

7: /cgn2_6/prodata/1/pubpaa/US11_NEW_PUB.pep.*

8: /cgn2_6/prodata/1/pubpaa/US60_NEW_PUB.pep.*

RESULT 1

US-10-522-789-2

Sequence 2, Application US/10522789

Publication No. US20050260732A1

GENERAL INFORMATION:

APPLICANT: TANABE SEIYAKU CO., LTD.

TITLE OF INVENTION: Three-dimensional structure of dipeptidyl peptidase IV

FILE REFERENCE: 03-039-PCT

CURRENT APPLICATION NUMBER: US/10/522,789

CURRENT FILING DATE: 2005-01-28

PRIOR APPLICATION NUMBER: US 60/398,761

PRIOR FILING DATE: 2002-07-29

NUMBER OF SEQ ID NOS: 2

SOFTWARE: PatentIn version 3.1

SEQ ID NO 2

LENGTH: 766

TYPE: PRT

ORGANISM: Homo sapiens

US-10-522-789-2

Query Match	24.5%	Score 987.5	DB 6	Length 766
Best Local Similarity	32.6%	Pred. No. 3.3e-67	Indels 83	Gaps 27
Matches	250	Conservative 131	Mismatches 303	
QY	11	TGEGSKRLTFNETVVKQAITPTSRVQWLSGAEDGSYVYAAEDGSLTIENIVTNESTRLI	70	Sequence 62, Appl
DB	36	TADSRKTYLTLD-YLKNTYRLKLYSLRWIS---DHEYLKQENNILVFNAYCGNSSVFLE	91	Sequence 60, Appl
QY	71	PADKIPTGKEAFNYMTHPDLSSVLWASNTKQVRHFFADYVQDVESLSKSVLMPDQOE-129		Sequence 46, Appl
DB	92	NSTDFEFGHSINDYSISPDGQFILLFVYVQWRHSYASY---DIYDLNKKRLQITEERI	148	Sequence 355, App
QY	130	-GDIQYAOHSYNGTITAFYRENDLYV--WDNGTVTTRITDDGGPD-MFHGVPDMWYEEEL	185	Sequence 4, Appl
DB	149	PNNQTQVTVSPVGHKLAYVNNNDIYVKBENLPSYRITWTGKEDIYNGITDMWYEEVF	208	Sequence 2, Appl
QY	186	GDRYALWFSPDGEBYLAYSFNETGVPTTYVQYMDNQEIAPAYFWELKIRYPKVSQNPRT	245	Sequence 173, App
DB	209	SAYSALWSPNGTFTFLAYAQFNDFEVLIEVSYFSDS---SLQPKTVRVPYPRAGAVNPT	265	Sequence 290, App
QY	246	VTLSLNI--ASKEVQKQPIDAFESTDLIG-----EVAWLTHHTTVAAKAFNRVQDQ-	297	Sequence 47, App
DB	266	VKFFVNTDLSLSSVTNATSIQITAPASMLIGHYLCDDVTWATQER--ISLQWLRRIQNY	323	Sequence 127, App
QY	298	-QKVAVDTASNKATVISDRD---GTDGWLNLNLSMKYIGPKPS-----DKDAYID	346	Sequence 29, Appl
DB	324	VMDICDYDESSGRWNCILVARQHIEMTTGM-----VGRFRPSEPHFTLDGNSFYKI	374	Sequence 2448, Ap

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	987.5	24.5	766	US-10-522-789-2	Sequence 2, Appl
2	985	24.4	762	US-11-116-939-13	Sequence 13, Appl
3	977	24.2	760	US-11-186-284-55	Sequence 55, Appl
4	243	6.0	657	US-11-179-977-1	Sequence 1, Appl
5	133.5	3.3	1424	US-11-102-476-4	Sequence 4, Appl
6	131	3.2	668	US-10-454-437-118	Sequence 118, App
7	123.5	3.1	621	US-10-467-657-1554	Sequence 1554, Ap
8	123.5	3.1	1045	US-11-113-424-54	Sequence 54, Appl
9	123.5	3.1	1094	US-10-821-234-1097	Sequence 1097, Ap
10	123.5	3.1	2376	US-11-096-051-4	Sequence 4, Appl
11	123.5	3.1	2715	US-11-096-051-2	Sequence 2, Appl
12	123.5	3.1	2721	US-11-096-051-10	Sequence 10, Appl
13	123.5	3.1	2725	US-11-096-051-8	Sequence 8, Appl
14	122.5	3.0	1259	US-10-995-561-625	Sequence 625, App
15	122.5	3.0	1286	US-10-995-561-628	Sequence 628, App
16	122.5	3.0	1341	US-10-995-561-621	Sequence 621, App
17	122.5	3.0	1348	US-10-995-561-624	Sequence 624, App
18	122.5	3.0	2355	US-10-995-561-623	Sequence 623, App
19	122.5	3.0	2355	US-10-995-561-627	Sequence 627, App
20	122.5	3.0	2384	US-10-821-234-1545	Sequence 1545, Ap
21	122.5	3.0	2386	US-10-995-561-626	Sequence 626, App
22	120.5	3.0	2715	US-11-113-424-51	Sequence 51, Appl
23	118.5	2.9	544	US-11-052-554A-358	Sequence 358, App
24	117	2.9	1315	US-10-995-561-630	Sequence 630, App
25	117	2.9	2296	US-10-995-561-633	Sequence 633, App

Qy 347 ISDSGWAHLVLPVSGGEPILTKGDWEVTSILSDOEROLVYLLSTQHHSTB--RHLY 404
Db 375 ISNEEGYRHCYFQIDKDKCTFTITKGTWEVIGIEALTSD--YLYIISNEYKMGPGGRNLY 432
Qy 405 SVSVSTFA-VTPL--VDDTVAAYWSASFSANSYVILTYGGPDVPYQELVYTTNSTKPLR 460
Db 433 KIQLSDYTKVTCLSCELNPERCQYVSFSKEAYYQLRCSGPGPLPLYTLHSSVNDKGLR 492
Qy 461 TITDNA---KYLEQIK-----DVALPNITYFELPSPGETLNMVQRLPPGFSPPDKKYP 511
Db 493 VLENSALDKMLQNVQPSKKLDFFILNETKFWQM-----ILPPHFDKSKKYP 542
Qy 512 LFTYGGGAEVTKRQWALNFKAYVADSLEEVVTVVNRGTFGKGRKRSVAVTRQLG 571
Db 543 LLDVYAGPSCQKADTVFR-LNWAYTLASTENI--IVASFDGRGSGYQGDKIMHAINRRLG 599
Qy 572 LLEAEDQIYAAQAAANIPWIDADHIGTWGSGVGLTSKLEKDSGAFTLGVTAPYSDW 631
Db 600 TFEVEDQLEAARQFSKMGFVNDKRIALMGWSYGGVTSVWLGSVGVKCGIAVAPVSRW 659
Qy 632 RFYDSMYTERYM--KTLSTNEEGYETSAV-RKTDGFKNVEGGFLIOHGTGDDNVHFQNSA 688
Db 660 EYDVSVYTERYMLPTPEDNLDRNSTVMSRAENFKQVE--YLLIHGTADDNVHFQNSA 717
Qy 689 ALVDLLMGDGVSPKLSHWTFDSDHGISYHGGGVFLYKQLARKLYQ 735
Db 718 QISKALVDVGVD---FQAMWYTTDEHGIASSTAQHIIYTHMSHFQ 761

RESULT 2

US-11-116-939-13
; Sequence 13, Application US/11116939
; Publication No. US20050265995A1
; GENERAL INFORMATION:
; APPLICANT: Stephen Tomlinson
; APPLICANT: Richard J. Quigg
; TITLE OF INVENTION: TISSUE TARGETED COMPLEMENT MODULATORS
; FILE REFERENCE: 19113.0115U2
; CURRENT APPLICATION NUMBER: US/11/116.939
; CURRENT FILING DATE: 2005-04-28
; PRIOR APPLICATION NUMBER: 60/565,907
; PRIOR FILING DATE: 2004-04-28
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 762
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence; note=synthetic
US-11-116-939-13

Query Match 24.4%; Score 985; DB 7; Length 762;
Best Local Similarity 32.9%; Pred. No. 5e-67;
Matches 245; Conservative 129; Mismatches 288; Indels 82; Gaps 26;

Qy 34 RSVWLGAEDGSVYAAEDGSLTIENIVTNESTRLLIPADKIPTKEAFNWIHPDLSSV 93
Db 54 KKLRLWIS---DHEVLYKQENILVNAEYGVSSVFLNSTDFEGHSINDYSISPDGQFI 110
Qy 94 LWASNHTKQYRHSFFADYVQDVESLKSVPMLPDQE--GDIQYQWSPVGNITAFVRND 151
Db 111 LLEYNVYQWRHSYTSY---DIYDLNKRQLITEERIPNNTQWVTSVPVGHKLAYVMND 167
Qy 152 LYV--WDNGVTTRITDDGPD-MFVGVDVWYEEIILDRVALWFSPPGEVLAFLSNFT 208
Db 168 IYVKIEPLRYSRITWTGKEDIYNGITDWYEEVFSAYSALMWSPNGTFLAYAFQNDT 227
Qy 209 GVPTTVQYVMDNQBIAPAYPWLKIRYPKVSQTNPTVTLISLNI--ASKEVQAPIDAF 266
Db 228 EVPLIEYSFYSD---SLQYPTKVRVPKAGAVNPTVKFVVVNTDSLSSVTNATSIQIT 284

Qy 267 ESTDLIIG-----EVAMLTDTHTTVAAKAFNRVQDQ--QKVVAVDTSANRATVLSRDD-- 317
Db 285 APASMLIGDHYLDCVDTWATQER--ISLQWLRRIQNYVMDCIDYDESSGRWNCVLAHQHI 342
Qy 318 --GTDGWLNDLLSMKYICPIKPS-----DKDAYIIDISDHSWAHLVLPVSGGEPIPL 369
Db 343 EMSTTGW-----VGRFPSEPHEFTLDGNSFYKIISNEEGYRHCYFQIDKDKCTFI 393
Qy 370 TKGDWEVTSILSDOEROLVYLLSTQHHSTB--RHLYSVSYSTFA-VTPL--VDDTVA 423
Db 394 TKGWEVIGIEALTSD--YLYIISNEYKMGPGGRNLYKIQLSDYTKVTCLSCELNPERCQ 451
Qy 424 YWSASFSANSYVILTYGGPDVPYQELVYTTNSTKPLATITDNA---KYLEQIK-----D 474
Db 452 YYSVSFSKEAYYQLRCSGPGPLPLYTLHSSVNDKGLRVLENSALDKMLQNVQPSKKLD 511
Qy 475 YALPNITYFELPSPGETLNMVQRLPPGFSPPDKKYPILFTPYGGGAEVTKRQWALNFK 534
Db 512 FIILNETKFWQM-----ILPPHFDKSKKYPILLDYYAGPSCQKADTVFR-LNWA 560
Qy 535 AYVADSLEEVVTVVNRGTFGKGRKRSVAVTRQLGLEAEDQIYAAQAAANIPWIDAD 594
Db 561 TYLASTENI--IVASFDGRGSGYQGDKIMHAINRRLGTFEVEDQLEAARQFSKMGFVNDK 618
Qy 595 HIGTWGSGVGLTSKLEKDSGAFTLGVTAPYSDWRFYDSMYTERYM--KTLSTNEEG 652
Db 619 RIATWGSYGGVTSVWLGSVGVKCGIAVAPVSRWEYVDSVYTERYMLPTPEDNLH 678
Qy 653 YETSAV-RKTDGFKNVEGGFLIOHGTGDDNVHFQNSAALVDLLMGDGVSPKLSHWTFD 711
Db 679 YRNSTVMSRAENFKQVE--YLLIHGTADDNVHFQNSAISKALVDVGVD---FQAMWYTD 733
Qy 712 SDHGISYHGGGVFLYKQLARKLYQ 735
Db 734 EDHGIIASSTAQHIIYTHMSHFQ 757

RESULT 3

US-11-186-284-55
; Sequence 55, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186.284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 55
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-11-186-284-55

Query Match 24.2%; Score 977; DB 7; Length 760;

Best Local Similarity 32.3%; Pred. No. 2e-66;
Matches 242; Conservative 140; Mismatches 293; Indels 74; Gaps 25;

QY 4 PRKHAPTEGSGKRLT-----FNETVVKQALPTSRVOWLSGAEDGSGTVYAAEDGSLTIE 59
DB 28 PSRVHNSSEMTWRLATLKDILNGTFSYKTFPP-----NWSIQE---YLHOSADNNIVLY 79
QY 60 NIVTNESRTLPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFPADYVVOQDVESL 119
DB 80 NIETQGSSTIIL-SNRTMKSVAASNYGLSPDRQFVYLEDSDYSLKWSYSTAYIYIDLNG 138
QY 120 KSVPLMPDQEGD-----IQYAWQSPVGNITAFVRENDLYVMD--NGTVTRITDDGGPD-M 171
DB 139 EPV-----RGNELPRPIQYLCWSVPVSGKLAIVYQNNIYKQRPDGPFPQITFNGRENKI 192
QY 172 FHGVPDWTYEEBILDRVALFSPDGEYLAYSFNETGVPTTVQYVYMDNQEIAPYWE 231
DB 193 FNGIPDWYEEBMLPTKALMWSNPKFLAYAEFNDKDIPIVAYSYYGDEQ-----YPR 247
QY 232 LKIRYPKVSOTNPTVTLN-----IASKEVKQAPIDA-PESTDLLIIEVAVLTDTHT 284
DB 248 INIPPKAGAKNPVRIIFIIDTTPAYGQEV---PVPAMTASSDYFSLTWTDER- 303
QY 285 TVAAKAFNRVQDQKVAVD-----TASNKATVISDRDGTGDLNLLSMKYIGPIKPS 338
DB 304 -VCLQWLXRVQNVSLICDPREDQWDCPKTQEHIEBSRTGWAGGFVSR--PVFSY 359
QY 339 DKDAYIIDISHGWAHLVLPVPSGGEPIPLTKGDWEVTSILSIDQERQLVYVLT--QH 396
DB 360 DAISYKIFSDKDGKHYIKDVENAIQITSGKWEAINPRVTQDS--LFYSSNEFEE 417
QY 397 HSTERHLSVSTF-----ATPLVDDTVAAWASFSANSYGYLITYGGDPVPOELY 451
DB 418 YPGRENIYRISGTSGPPSKKCTCHLKERCOYITASFSYAKYALVCYGGIPISTUH 477
QY 452 TTNSKPLRTIDNAKLEQIKDVALPNITYFELPLPSGETLNNMQLPPGSPDKKYP 511
DB 478 DORTQOEIKILEENKELENALKNIQLPKEIKKLEVE-ITLWYKMLPPQDPSKYP 536
QY 512 LFTYGGPGAQVTRKQWALNFKAYVASDSSELYVTWVDNRGTGFKGRKFRSAVTRQLG 571
DB 537 LIQVYGGPCQSRSVF-AVNWISYLASKEGM--VIALVDRGTAFQGDKLLYAVYRKIG 593
QY 572 LLEAEDQIYAAQAAANIPWIDADHIGWMSFGGYLTSKVEKDSGAFTLGVITAPVSDW 631
DB 594 VYEVEDQITAVRKFTIEMGFIDKRIALWMSYGGIVSSIALASGTLFCGIAVAPVSW 653
QY 632 RFVDSMYTERYM--KTLSTNEEGYETSAY-RKTDGFKNVEGGFLIOHGTGDDNVHFQNSA 688
DB 654 EYASVYTERFMGLPTKDDNLEHYKNSYTWARAEYFRVD--YLLHGTADDNVHFQNSA 711
QY 689 ALVDDLMDGVSPEKLHSGWFTSDSHGIS 717
DB 712 QIAKALVNAQVD---FOAMWYSQNHGLS 737

Query Match 6.0%; Score 243; DB 7; Length 657;
Best Local Similarity 20.9%; Pred. No. 6.6e-11;
Matches 146; Conservative 113; Mismatches 333; Indels 106; Gaps 26;

QY 67 RTLIPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFPADYVVOQDVESLKSVPMLP 126
DB 2 KKLITADDITAIIVSVTDPOYAPDGTAAAYVKSQVQNEKDSYTSNIWIYETKTGGSVPTW 61
QY 127 DQEGDIQYVAQWSPVGNITAFV--REND---LYVM--DNGTVTRITDDGGPDPMFHGVPDWI 179
DB 62 GEKSTD-PRNSPDGRTLAFISDRGDAALYIINSTEAGEARKLTD-----IP--- 108
QY 180 YEEIEILDRYALWFSPDGE-YLAYLSFNETGVPTTVQYVYMDNQEIAPYWEKIRYDK 238
DB 109 -----YGVSKPLW-SPDGESILVTISLGESESIDDRKTEQDSYEPVEVQGLSYK-RD 161
QY 239 VSOQNTPTVTLNLTASKEVQAPIDAFESTDILLIIEVAVLTDTHTVAAKAFNRVQDQ 298
DB 162 GLTRGAYAQVLVSVKSGEMKE-----LTSKADHGDPAFSPDGKWLVPFSAFLTETDDAS 216
QY 299 K---VAVDVTASNKATVISDRDGTGDLNLLSMKYIGPIKPSDKD-----AYYIDI 347
DB 217 KPHDVIMLSLSDGLKQVTPHRSFGSSFSFDPGRYLA-LLGNEKEYKNATLSKAWLYDI 275
QY 348 SD-----HSGWAHLVLPVPSGGEPIPLTKGDWEVTSILSIDQERQLVYVLTQ 395
DB 276 EGRLTCLTEMLDVHLADALIGDSLIGGAEQRPITWKDSQGFYVIGTDQSGTGYYI--- 332
QY 396 HSTERHLSVSYSTFAVTPVDDTVAAWASFSANSYGYLITYGGDPVPOQELVYVLTNS 455
DB 333 --SIEGLVYIRLEKEYIN-----SFLSPDEQHFASVTKPRDP-SELYSIFL 378
QY 456 TKPLRTITD-NAKLEQIKDVALPNITYFELPLPSGETLNNMQLPPGSPDKKYPILFT 514
DB 379 GQEEKQLTGANDKF---VREHTISIPETIYATEQGVWNGWMLRPAQNEGETTYFLILN 435
QY 515 PYGGPG---AQEVTKRQWALNFKAYVASDSSELYVTWVDNRGTGFKGRKFRSAVTRQLG 571
DB 436 IHGGPHMYGHTYFHEFQVLAAGYAVV-----YINPRGSHGYGQEFVNAVVRGDY 486
QY 572 LLEAEDQIYAAQAA-ANIPWIDADHIGWMSFGGYLTSKVEKDSGAFTLGVITAPVSD 630
DB 487 GKDYDDVMQAVDEAIKRDPIHIDPKRLGVGTGSGYGGFMTNWIYVQ--TNRFKAAVTRQ 545
QY 631 RFVDSMYTERYMKT-----LSTNEEGYETSAYRKTGDKFNVEGGFLIOHGTGDDNV 682
DB 546 WISFHVSDIGYFTFDWQLEHDMFEDTEKLWDRSPKLYA---ANVETPLLILHGERDDRC 602
QY 683 HFQNSAALVDDLMDGVSPEKLHSGWFTSDSHGISYHG 720
DB 603 PIEQAEQLFIALKWKGTETKLV---FPNASHNLSRTG 637

RESULT 5
US-11-102-476-4
; Sequence 4, Application US/11102476
; Publication No. US20050271680A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Yung-Fu
; APPLICANT: Palaniappan, Raghavan U.M.
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: Novel Immunogenic Proteins of Leptospira
; FILE REFERENCE: 1153.080US1
; CURRENT APPLICATION NUMBER: US/11/102,476
; PRIOR FILING DATE: 2005-04-08
; PRIOR APPLICATION NUMBER: PCT/US2003/32385
; PRIOR FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: US 60/417721
; PRIOR FILING DATE: 2002-10-10
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4

Best Local Similarity 32.3%; Pred. No. 2e-66;
Matches 242; Conservative 140; Mismatches 293; Indels 74; Gaps 25;

QY 4 PRKHAPTEGSGKRLT-----FNETVVKQALPTSRVOWLSGAEDGSGTVYAAEDGSLTIE 59
DB 28 PSRVHNSSEMTWRLATLKDILNGTFSYKTFPP-----NWSIQE---YLHOSADNNIVLY 79
QY 60 NIVTNESRTLPADKIPGKAFNWIHPDLSSVLWASNHTKQYRHSFPADYVVOQDVESL 119
DB 80 NIETQGSSTIIL-SNRTMKSVAASNYGLSPDRQFVYLEDSDYSLKWSYSTAYIYIDLNG 138
QY 120 KSVPLMPDQEGD-----IQYAWQSPVGNITAFVRENDLYVMD--NGTVTRITDDGGPD-M 171
DB 139 EPV-----RGNELPRPIQYLCWSVPVSGKLAIVYQNNIYKQRPDGPFPQITFNGRENKI 192
QY 172 FHGVPDWTYEEBILDRVALFSPDGEYLAYSFNETGVPTTVQYVYMDNQEIAPYWE 231
DB 193 FNGIPDWYEEBMLPTKALMWSNPKFLAYAEFNDKDIPIVAYSYYGDEQ-----YPR 247
QY 232 LKIRYPKVSOTNPTVTLN-----IASKEVKQAPIDA-PESTDLLIIEVAVLTDTHT 284
DB 248 INIPPKAGAKNPVRIIFIIDTTPAYGQEV---PVPAMTASSDYFSLTWTDER- 303
QY 285 TVAAKAFNRVQDQKVAVD-----TASNKATVISDRDGTGDLNLLSMKYIGPIKPS 338
DB 304 -VCLQWLXRVQNVSLICDPREDQWDCPKTQEHIEBSRTGWAGGFVSR--PVFSY 359
QY 339 DKDAYIIDISHGWAHLVLPVPSGGEPIPLTKGDWEVTSILSIDQERQLVYVLT--QH 396
DB 360 DAISYKIFSDKDGKHYIKDVENAIQITSGKWEAINPRVTQDS--LFYSSNEFEE 417
QY 397 HSTERHLSVSTF-----ATPLVDDTVAAWASFSANSYGYLITYGGDPVPOELY 451
DB 418 YPGRENIYRISGTSGPPSKKCTCHLKERCOYITASFSYAKYALVCYGGIPISTUH 477
QY 452 TTNSKPLRTIDNAKLEQIKDVALPNITYFELPLPSGETLNNMQLPPGSPDKKYP 511
DB 478 DORTQOEIKILEENKELENALKNIQLPKEIKKLEVE-ITLWYKMLPPQDPSKYP 536
QY 512 LFTYGGPGAQVTRKQWALNFKAYVASDSSELYVTWVDNRGTGFKGRKFRSAVTRQLG 571
DB 537 LIQVYGGPCQSRSVF-AVNWISYLASKEGM--VIALVDRGTAFQGDKLLYAVYRKIG 593
QY 572 LLEAEDQIYAAQAAANIPWIDADHIGWMSFGGYLTSKVEKDSGAFTLGVITAPVSDW 631
DB 594 VYEVEDQITAVRKFTIEMGFIDKRIALWMSYGGIVSSIALASGTLFCGIAVAPVSW 653
QY 632 RFVDSMYTERYM--KTLSTNEEGYETSAY-RKTDGFKNVEGGFLIOHGTGDDNVHFQNSA 688
DB 654 EYASVYTERFMGLPTKDDNLEHYKNSYTWARAEYFRVD--YLLHGTADDNVHFQNSA 711
QY 689 ALVDDLMDGVSPEKLHSGWFTSDSHGIS 717
DB 712 QIAKALVNAQVD---FOAMWYSQNHGLS 737

US-11-179-977-1
; Sequence 1, Application US/11179977
; Publication No. US20050249789A1
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GC511-PCT
; CURRENT APPLICATION NUMBER: US/11/179,977
; CURRENT FILING DATE: 2005-07-12
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
US-11-179-977-1

```
; LENGTH: 1424
; TYPE: PRT
; ORGANISM: Leptospira interrogans
US-11-102-476-4

Query Match      3.3%; Score 133.5; DB 7; Length 1424;
Best Local Similarity 19.0%; Pred. No. 0.04;
Matches 164; Conservative 103; Mismatches 307; Indels 291; Gaps 38;

Qy 28 AITPTSRVQWLS-----GAEDGSVVAAEDGSLTIENIVNESRTL-----IPADKI 75
Db 92 SIVPDSQSVTTIQNRVGRGIASGSIKAEVNGLYSEQKIIVTPTATLNSIQVTSLESGIL 151
Qy 76 PTGRE-----AFNYWIHPDLSS---VLWASNH-----TKQYRHSF 107
Db 152 PKGTNRQFSAIGFSDGSHQDISNDPLIVWSSNPDLVQVDDSLASGLNLTGTAHIRASF 211
Qy 108 -----PADYVQDVESLKSVP LMP-----DQEGDIQYAO--WSPVG 141
Db 212 QSKQGAEMTVGDAVLSQIQVTSNPNPIPLGKKQKLIATGIYSDNSNRDISSSVIMWSSN 271
Qy 142 NTIAFVREND--LYVMDNGTVTRITDDGCPDMFHGVDPDWIYEEELGD-----RYAL 191
Db 272 STIANIQNGILETADGIVTISAS-----SENIIGSKLIVTPAALVSI 316
Qy 192 WFSPDGEYLA---YLSFNETGVPTTYVQYVMDNQEIAPAYPWE-----LKIRYPK---- 238
Db 317 SVSPNTSVAKGLOENFKATGIFTN-----SNSDITDQVTDWSSNTDILSISNASDSHG 371
Qy 239 ----VSQTNPTVTLNLNIASK---EVKQAPIDAFESTDLIIGEVAMLTHTHTTVAAKAP 291
Db 372 LASTLNQGNKVTASIGGIQGSTDPVTQAAITSIKQVSPVLPSTAKGLTKRFTAGIFTD 431
Qy 292 NRVDQKQV-----AVDTASN-----KATVISPRDCT-----DGMW-----DNL 326
Db 432 NSKKDITDQVTDWSSSAIVSVNSLDDNKGKGAHAAGDTTITATLKGKVGKTLVTPVPAV 491
Qy 327 LSMKYIGIPKPSDKDAYYIDISDHSWAHLVLPFVSGGEPIPLTKG---DWEVTSILSID 383
Db 492 LTSIQINPVNPS-----LAKGLTKQFSATGIYSDN 521
Qy 384 QERQLVYLLSTQHHSSTERHLYSVSVSTPAVPLVDDTVAATWSASFSANSYYILTLYGGP 443
Db 522 SNKDI-----TSAVTFSSSDSIATISNAQKNQGNAYGAATGAT 560
Qy 444 DVPQELYTNTKPLRTITDNAKVLKLEOI KDYALPNITYPELPLPSGETLNVQMQLRPGF 503
Db 561 DI---KATFGKVSSPVSTLSVTAKLVEIQ-----587
Qy 504 SPDKKYPILFTPYGGGQAQVTKRWQALNFKAYVASDSELEYVTWTVDNRG-----TGPKG 559
Db 588 -----ITPAASAKAKGLTERFKATGIFTDSSNSDITNQVTWSSNTDIAEITNTSG 638
Qy 560 RKFRSAVTQQLGLBAEDQIVAAQQAANI PMIDADHIGWSPGGYLTISKVLEKDSGAF 619
Db 639 SK--GITNT--LTPGSSEISAA-----LGSIKSSKVLKVTAPAQ 673
Qy 620 TLGVITAPVSD-----WRFYDS--MYTERYMKTLSTNEEGVETSAVRKTDGFKNVGGF 671
Db 674 LISIAVTPIINSAVKGLRIQFKATGTYTDHVSQDDV--TALATWSSSNPRK--AMNVNVTGS 730
Qy 672 LIQHTGDDNVHFQNSAALVLLMDD---GVSPKLSHQWFETSDSHGISYHGGGVFLYKQ 728
Db 731 VTTVATGNTNI-----KATIDSIGSSVLNVTPALLTISIEITPTINSIT-HG-----LTKQ 780
Qy 729 L-ARKLYOEKXNRQTVLHMWTKKD 752
Db 781 FKATGIFSDKSTQNLTLQVLTWISSD 805

RESULT 6
US-10-454-437-118
; Sequence 118, Application US/10454437

Publication No. US20050277115A1
; GENERAL INFORMATION:
; APPLICANT: Pompejus, Markus
; APPLICANT: Krogger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberhauser, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN HOMEOSTASIS AND ADAPTATION
; FILE REFERENCE: BGI-128CPCN
; CURRENT APPLICATION NUMBER: US/10/454,437
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/141031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: DE 19931636.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19932125.6
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932129.9
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: DE 19932226.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932920.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932922.2
; PRIOR FILING DATE: 1999-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 442
; SEQ ID NO 118
; LENGTH: 668
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-454-437-118

Query Match      3.2%; Score 131; DB 6; Length 668;
Best Local Similarity 17.8%; Pred. No. 0.021;
Matches 138; Conservative 100; Mismatches 270; Indels 268; Gaps 34;

Qy 78 GKEAFNYWIHPDLSSVLWASNHKQYRHSFPADYVQDVESLKSVP LMPDQEGDIQYAO 137
Db 59 GEKLYNFRDAQHPRGVWRITTLSEYE---SDQPEWDV--LIDVDALAEDEG----- 105
Qy 138 SPVGNITAFVRENDLYVMDNGTVTRITDDGCPDMFHGVDPDWIYEEELGDRYALWFSPDG 197
Db 106 -----EN--WVW--KGAVVR-----SPEF 120
Qy 198 EYLAVLSFNETGVPTTYVQYVMDNQEIAPAYPWEKIRYKVSQTNPTVTLNLNIASK 257
Db 121 D-RALVFRSGGADATVIREF-----DLATAAFVDDSPFE 154
Qy 258 VKQAPIDAFESTDLIIGEVAMLTHTTVAAKAFNRVQDQKVVAVDTASNAKATVISORD 317
Db 155 LKEAKSD-----VTWV-DLDTL-----LVGTDGEGSLT----- 182
Qy 318 GTDGLWNLNLSMKYIGIPKPSDK--DAYYIDISDHSWAHLVLPFVSGGEPIPLTKGDWE 375
Db 183 -DSGYPARVLTVTKRGTPLEQAELFPEGSRQDVATHA-WR-----DSTPG-FE 226
Qy 376 VTSILSIDQERQLVYLLSTQHHSSTERHLYSVSYST-----FAVTPLVDDTVAAYW 425
Db 227 RTFV-----SRSLDFYNSETSLETGGLVLDVPTDCDVIVKKQWIFVSPRTD-----F 275
Qy 426 SASFSANSQYIYIL---TYGGDPVYPQELYTTNNTSKPLRTI---TONAKVLEKIQKDAL--- 477
Db 276 AGIPAGGLGVLLKKEFLGEGGRD--FQPVFTPTTESTSLQGLATTNKNFLVTLNNVSTEIV 333
Qy 478 -----PNITYFELPLPSGETLNVW-----QRLPFGFSPD 506
```

Db 334 TVPLNDPTEHEHIDLPBHVAHVATSPDGDGEIHWQAASFEAPTILLRAELPGALEAV 393
Qy 507 KYPILFTPGGPGAGQVETKRW-----QALNPKAY-----VA 538
Db 394 KKAPOFENAG-----QETROHWATSADGKIPYFITGAFEBEPQNTLVHAYGGFEVSLTP 449
Qy 539 SDSELEYVTW-----TVDNRGTFKGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPW 590
Db 450 SHSPTRGTAWLEKGYFFVEANLRGGEGEPHWSQATKLNRMKWEDHRAVLADLVERGY 509
Qy 591 IDADHIGWMSFGYLSKVLKSDGAPFTLGVITAPVSDW-RFYDSMYTERYMKTLSTN 649
Db 510 ATPEQIAIRGGSGGLLTSGALTQYPEAFGAAVQVPLADMLRYHTVSAGASWMAEYGNP 569
Qy 650 BEGYETSARKTDGFKNVEG-----GFLIOHGTGDDNVHFNQNSAALVDLLMGDGVSP 702
Db 570 DDEPERAVIEQYSPQAVGVGVEKRYPPALVTTSTRDRVHPAHARLFAQALLDGAQAVD 629
Qy 703 KLHSQWFTSDSHGISHYGGGVLYKQARKLYQEKNRQT---QVLHMQHTKKDLEE 755
Db 630 -----YENTEGG---HAGAA-----DNKQTAFFVESLIYTWIEKTLQ 664

RESULT 7

US-10-467-657-1554
; Sequence 1554, Application US/10467657
; Publication NO. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 1554
; LENGTH: 621
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-1554

Query Match 3.1%; Score 123.5; DB 6; Length 621;
Best Local Similarity 20.1%; Pred. No. 0.069;
Matches 134; Conservative 83; Mismatches 226; Indels 225; Gaps 37;
Qy 129 EGDIOYAQWSPVG-NTIAFVRENDLYV---WNGTWTTRITDGGPDMFHGVPDWEYEEI 184
Db 104 EGGFHF---PAGKHVSWRDENVVWCPAWE---RQLTESQ-----YPREV 144
Qy 185 LGDRYALWSPDGEYLAISFNETGYPTTVQ-----YYMDNQEITAPVPWELKIR 235
Db 145 -----WLVERGK-----SFEES-LPAYQIDKGMVNAVRYLDPQ----- 178
Qy 236 YPKVSTQNTVTLSLNLASKEVKQAPIDAFESTDLIIGEVAMLTDTHTTVA-----AKAF 291
Db 179 -----GSPIDLEASD-----GFYTKYTLQVSSEGKAPL 208
Qy 292 NRVDQOKVAVDVTASNKATVISDRDGTGWLDN---LLSMKYIGPIKPSDKXAYIIDSD 349
Db 209 NLPNDQDVV-----GYLAGHLLTLR-----KWHRANQS- 238
Qy 350 HSGWAHLYLFPVGGGEPILTKGDWEVTSILSDQERQLVYVLSQHHSTERHLYSVS 409
Db 239 -----YPSGALVAVKLNRGELGAAQLLFPAPDETALESVET-----TKR----- 277
Qy 410 TFAVTPLVDDTVAAVWSAFSFA-----SGYIILT---YGGPDVPVQELVTTNS 455

Db 278 -FVVASLLENVQGRKAWRFADSKWQBAELPHLPFGALEMTDQPMGG-DVVT--LAASDF 333
Qy 456 TKPLRTITDNKVLK-----QIKDYALPNITYFELPLPSGETLNVQMRLP-----PGFS 504
Db 334 TTPLTLFALDNLNVELTVMRLQPOQFVSDGIEVRQFMAVSSDG-----ERIPYFHVGNAA 389
Qy 505 PDKKYPILFTPYGGGAQE-----VTKRQALNFKAYVASDSSELEYVTWTVDRNGTGF 557
Db 390 PDT--PTLVAYAGGFGIPELPHYLGVSQVKYMLE-EGNAFVLANI-----RGGE 435
Qy 558 KGRKFRSAVTRQLGLLEAEDQIYAAQQAANIPWIDADHIGWMSFGGYLTSKVLKSDG 617
Db 436 FGPRHQAQGISKHKSVDDLLAVRDJSEKMSGPKHIGLOGGNGGLUITAAAFVREPO 495
Qy 618 AFLTGLVITAPVSDWRFY-----DSMYTBRYMKTLSTNEEGYETSAR--KTDGPKNVEGG 670
Db 496 SIGALVCEVPLTDMTRYPLLSAGSSWTDEY-----GNPKYEAACKRRLGELSPYHNLSDG 550
Qy 671 F-----LIHGTGDDNVHFNQNSAALVDLLMGDGVSPKLSQWFTSDSHGISHYGGGVFL 725
Db 551 IDYPPALITTSLSDDRHPAHALKFYAKLR--ETSPQ-----SWLYSPDGG-GHTGNGT-- 601
Qy 726 YKQLARKL 733
Db 602 QRESADKL 609

RESULT 8

US-11-113-424-54
; Sequence 54, Application US/11113424
; Publication NO. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 54
; LENGTH: 1045
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-54

Query Match 3.1%; Score 123.5; DB 7; Length 1045;
Best Local Similarity 20.0%; Pred. No. 0.15;
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;
Qy 60 NIWT--NESRTIIPADKIPTGKEA-ENYWIHPDLSSVLWASNHTKQYRHSFFADYVQDV 116
Db 269 SIITDNEGLLLQTAFLGTSSRVLFKYRRQTFLSILYDST-----RVSF---TYDETA 320
Qy 117 ESKSVPLMPDQEG---DIQYAQWSPVGNITAF-----VRENDLYVWDNGTGTTRITDD 166

```
Db 321 GVLKTNLQSD--GFICTIRYQIGPLIDRQIFRSEDGMVNFARFDSYDNS--FRVTSM 376
Qy 167 GGPDMFHGVPDWIYEIEELGDRYALWFSPOGEY-LAYLSFNE---TCGPTVYT-----214
Db 377 QGVNETPLPIDLYQFDDISGKVEQF---GKFGVIYDINQIISTAVMTYTKHFDAGR 432
Qy 215 ---VQYMDNQEIAPAYPWPWELKIRYPKVSQTNPTVTLNLNIAKSKVQAP-----262
Db 433 IKEIQY-----EIFRSLMYWITIYDNMGR-----VTGREIKIGFPANTTKVAY 476
Qy 263 ---IDAFESTDLIIGEVAWLTDTHTTVAAKAFNRVQDQKVAVDTASNKATVISD---R 316
Db 477 EYVDGQLQTVYLNKIMRWYNYDLNGLNHLNPSNARLTPLRYDLRDRITRLGDVQYR 536
Qy 317 DGTGWL-----DNLISMKYIGPIKPSDKDAYIIDISDHSGMAHLXLPVSGG 364
Db 537 LDEGFLRQGRTEIFEYSSKGLLTRVY-----SKSGWTVIYRYDGLGR 580
Qy 365 EPIPLTKGDMWVTSILSDQERQLVYVYLSLTOHHSTERHLYSVSYSTFAVTPLVDDTVAAY 424
Db 581 -----RVSKTSLGQHLQF-FYADLTYPTRITHVY--NHSSEITSLYDLOHGL 627
Qy 425 WSASFANSYGYIITYGGPDVPYQELYTNTSTKPLRTITDNKVLQEIQKVALPNITYFE 484
Db 628 FAMEISSGDEFYIA-----SDNTGTPLAVFSSNGLMLKQIQYTAYGEI-YFD 673
Qy 485 LPLSGETLNVQRLPPGFS-----PDKKYPILFTPYGGPGAEVTKR--WQ 529
Db 674 -----SNIDFQLVIGFHGGLYDPLTKLIHFGERYDILAGRWTTPDI-EIWKRIKGD 724
Qy 530 ALNFKAYVADSE-----LEVYT---WTV-----DNRGTGFKRKFERSAVTRQLG 571
Db 725 PAPNLNMFNRNNPASKIHDVKDIITDVSMLVTFGPHLHNAIPGFPVPKF-----D 776
Qy 572 LLEAEDQIYAAQQAANIPWIDADHIGI-----PDKKYPILFTPYGGPGAEVTKR--W 599
Db 777 LTPESYELVKSQWDDIPPI-----FGVQQQVARQAKAFSLGKMAEVQVSRRRAGQSW 832
Qy 600 GWSFGGYLTSK-----VLEKDSGAPTLGVI-----TAPVSDWRFY-----D 635
Db 833 LW-----FATVKSLLIGKGVMLAVSQGRVQTNVLNINANEDCIKVAALNNAFYLENLHFTIE 888
Qy 636 SMYTERYMKTLSTNEEGVETSAAVKTGDKFNVEGGFLIQHGTGDDNFHONSAAALV-----691
Db 889 GKOTHYFIKT--TTPES-DLGTLRLTSGRKALENGI-----NVTVSQSTTVVNGRT 936
Qy 692 ----DLMGDGVSPKELHSQWFTSDHGISYHGGVFLYKOLARKLYQEKVRQTOVLHQ 747
Db 937 RRFADVENQFGAL--ALHVRV-----GMTLDEEKARILEQARQ---ALARA 978
Qy 748 WTK-----KDL EE 755
Db 979 WAREQQRVRDGE 991
```

RESULT 9

```
US-10-821-234-1097
; Sequence 1097, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
```

```
; SEQ ID NO 1097
; LENGTH: 1094
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1097
```

Query Match 3.1%; Score 123.5; DB 6; Length 1094;

Best Local Similarity 20.0%; Pred. No. 0.16;

Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;

```
Qy 60 NIVT--NESRTLPAADKIPGKEA-FNYMIHPDLSVLMASNHTKQYRHSFFADYVQDV 116
Db 318 SIITDYNEGLLQAPLGTGSRRLFKYRRQTLSEILYDST-----RVSF---TYDETA 369
Qy 117 ESLKSVPLMPDQEG---DIQYAQMSVPGNTIAP-----VRENDLYWONGTVTRTDD 166
Db 370 GVLKTNLQSD--GFICTIRYQIGPLIDRQIFRSEDGMVNFARFDSYDNS--FRVTSM 425
Qy 167 GGPDMFHGVPDWIYEIEELGDRYALWFSPOGEY-LAYLSFNE---TCGPTVYT-----214
Db 426 QGVNETPLPIDLYQFDDISGKVEQF---GKFGVIYDINQIISTAVMTYTKHFDAGR 481
Qy 215 ---VQYMDNQEIAPAYPWPWELKIRYPKVSQTNPTVTLNLNIAKSKVQAP-----262
Db 482 IKEIQY-----EIFRSLMYWITIYDNMGR-----VTGREIKIGFPANTTKVAY 525
Qy 263 ---IDAFESTDLIIGEVAWLTDTHTTVAAKAFNRVQDQKVAVDTASNKATVISD---R 316
Db 526 EYVDGQLQTVYLNKIMRWYNYDLNGLNHLNPSNARLTPLRYDLRDRITRLGDVQYR 585
Qy 317 DGTGWL-----DNLISMKYIGPIKPSDKDAYIIDISDHSGMAHLXLPVSGG 364
Db 586 LDEGFLRQGRTEIFEYSSKGLLTRVY-----SKSGWTVIYRYDGLGR 629
Qy 365 EPIPLTKGDMWVTSILSDQERQLVYVYLSLTOHHSTERHLYSVSYSTFAVTPLVDDTVAAY 424
Db 630 -----RVSKTSLGQHLQF-FYADLTYPTRITHVY--NHSSEITSLYDLOHGL 676
Qy 425 WSASFANSYGYIITYGGPDVPYQELYTNTSTKPLRTITDNKVLQEIQKVALPNITYFE 484
Db 677 FAMEISSGDEFYIA-----SDNTGTPLAVFSSNGLMLKQIQYTAYGEI-YFD 722
Qy 485 LPLSGETLNVQRLPPGFS-----PDKKYPILFTPYGGPGAEVTKR--WQ 529
Db 723 -----SNIDFQLVIGFHGGLYDPLTKLIHFGERYDILAGRWTTPDI-EIWKRIKGD 773
Qy 530 ALNFKAYVADSE-----LEVYT---WTV-----DNRGTGFKRKFERSAVTRQLG 571
Db 774 PAPNLNMFNRNNPASKIHDVKDIITDVSMLVTFGPHLHNAIPGFPVPKF-----D 825
Qy 572 LLEAEDQIYAAQQAANIPWIDADHIGI-----PDKKYPILFTPYGGPGAEVTKR--W 599
Db 826 LTPESYELVKSQWDDIPPI-----FGVQQQVARQAKAFSLGKMAEVQVSRRRAGQSW 881
Qy 600 GWSFGGYLTSK-----VLEKDSGAPTLGVI-----TAPVSDWRFY-----D 635
Db 882 LW-----FATVKSLLIGKGVMLAVSQGRVQTNVLNINANEDCIKVAALNNAFYLENLHFTIE 937
Qy 636 SMYTERYMKTLSTNEEGVETSAAVKTGDKFNVEGGFLIQHGTGDDNFHONSAAALV-----691
Db 938 GKOTHYFIKT--TTPES-DLGTLRLTSGRKALENGI-----NVTVSQSTTVVNGRT 985
Qy 692 ----DLMGDGVSPKELHSQWFTSDHGISYHGGVFLYKOLARKLYQEKVRQTOVLHQ 747
Db 986 RRFADVENQFGAL--ALHVRV-----GMTLDEEKARILEQARQ---ALARA 1027
Qy 748 WTK-----KDL EE 755
Db 1028 WAREQQRVRDGE 1040
```

RESULT 10

US-11-096-051-4

```
; Sequence 4, Application US/11096051
; Publication No. US20050244868A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacLachlan, Timothy K
; APPLICANT: Rastelli, Luca
; APPLICANT: Vernet, Corine
; APPLICANT: Ettenberg, Seth
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use
; FILE REFERENCE: Attorney Docket No. Cura 967
; CURRENT APPLICATION NUMBER: US/11/096,051
; CURRENT FILING DATE: 2005-03-30
; PRIOR FILING DATE: 2001-12-31
; PRIOR FILING DATE: 2003-06-04
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: 60/557,978
; PRIOR APPLICATION NUMBER: 60/557,978
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 4
; LENGTH: 2376
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-11-096-051-4

Query Match      3.1%; Score 123.5; DB 7; Length 2376;
Best Local Similarity 20.0%; Pred. No. 0.47;
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;

QY      60 NIWT--NESRTLIPADKIPTGKEA-FNYWIHPDLSSVLWASNHTKQYRHSFFADYYVQDV 116
DB      1600 SIITDYNESGLLQTAFLGTSRRVLFKYRRQTFELSEILDST-----RVSF---TYDETA 1651

QY      117 ESKSVPLMPDQEG---DIQYQMSVGNNTIAF-----VRENDLYVWDNGTTRITDD 166
DB      1652 GVLKTNVQLSD--GFICTIRYQIGPLIDRQIFRSEDGMVNAFYSYDNS--FRVTSM 1707

QY      167 GGPDMFHGVDPDMEIEELGDRVALWFSPDGEY-LAYLSFNE---TGVPYTYT----- 214
DB      1708 QGVINETPLDLYQFDDISGKVEQF-----GKFGVIYDINOIISTAVMTYTKHFDAGH 1763

QY      215 ---VQYMDNOEIAPAYPMELKIRYKPSQTNPTVTLSSLNIAKSEVKQAP----- 262
DB      1764 IKEIQY-----ELFRSLMYWITIYDYMGR-----VTKREIKIGFPANTTKYAY 1807

QY      263 ----IDAFESTDLIGEVAMLTDTHTTAAKAFNRVQDQKVAADVTASNKATVISD---R 316
DB      1808 EYDVEDQLOQTVYLNKIMWRNYDNLGNLHLPNSNSARLTPLRYDLDRITRLGQVQYR 1867

QY      317 DGTGHWL-----DNLISMKYIGIPKPSDKDAYVIDISDHSQWAHLVLPFVSGG 364
DB      1868 LDEDFGLRQRGTEIFBYSSKGLLTRVY-----SKSGGWTVIYRYDGLGR 1911

QY      365 EPIPLTKGDWEVTSILSDQERQLVYLSLTHSTHSLYVSYSVAVTAVTPLVDDTVAAY 424
DB      1912 -----RVSSKSLGHLQF-PYADLTYPTRITHVY--NHSSEITSLYDLOCHL 1958

QY      425 WSASFSGANGSYITLYTGGPDVPYQELTYNTSTKPLRTITDNKVLQIKDYALPNTIYFE 484
DB      1959 FAMEISSGDEFYIA-----SDNTGTPFLAVFSSNGLMLKIQIQTAYGEI-YFD 2004

QY      485 LPLPSGETLNMQRLLPPGFS-----PDKKYPIILFTPYGGPGQAQVTKR--WQ 529
DB      2005 -----SNIDFQLVIGFHGGYDPLTKLIHFGERDYDILAGRWTTTPII-ELWKRIKGD 2055

QY      530 ALNFKAVVADSE-----LEVVT-----WTV-----DNRGTGFKGRFRSAVTRQLG 571
DB      2056 PAPFNLYMFRNNPASKIHVDKDYITDVSNSWLVTFGFHLHNAIPGFPVPEF-----D 2107

QY      572 LLEAEQDIYAAQQAANIPWTDADHIGI-----FGVQQQVARQAKAFLSLGKVAEQVSSRRRAGGAQSW 599
DB      2108 LTPESYELVKSQQWDDIPPI-----FGVQQQVARQAKAFLSLGKVAEQVSSRRRAGGAQSW 2163
```

```
QY      600 GWSFGGYLTSK-----VLEKDSGAFTLGV-----TAPVSDMRFY-----D 635
DB      2164 LW-----FATVKSLIGKVMALVVSQGRVQTNVLNINIANEDCIKVAALVNNAFYLENLHFTIE 2219

QY      636 SMYTERYMKTLSTNEEGYETSAVRKTGDKFNKVEGGFLQHGTDGDDNVHFQNSAALV---- 691
DB      2220 GKOTHYFIKT--TTPES--DLGTLRLTSGRKALENGI-----NVTVSQSTTVVNGRT 2267

QY      692 ----DLLMGDGVSPKHLHSQWFTSDHGISVHGGGVFLYKQLARKLYQEKNRQTOVLMHQ 747
DB      2268 RRFADVEMQFGAL--ALHVRV-----GWLDEEKARILEQARQR---ALARA 2309

QY      748 WTK-----KDLSE 755
DB      2310 WAREQORVRDGE 2322

RESULT 11
US-11-096-051-2
; Sequence 2, Application US/11096051
; Publication No. US20050244868A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacLachlan, Timothy K
; APPLICANT: Rastelli, Luca
; APPLICANT: Vernet, Corine
; APPLICANT: Ettenberg, Seth
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use
; FILE REFERENCE: Attorney Docket No. Cura 967
; CURRENT APPLICATION NUMBER: US/11/096,051
; CURRENT FILING DATE: 2005-03-30
; PRIOR FILING DATE: 10/038,854
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 10/455,772
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: 60/557,978
; PRIOR FILING DATE: 2004-03-30
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 2
; LENGTH: 2715
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-11-096-051-2
```

```
Query Match      3.1%; Score 123.5; DB 7; Length 2715;
Best Local Similarity 20.0%; Pred. No. 0.57;
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;

QY      60 NIWT--NESRTLIPADKIPTGKEA-FNYWIHPDLSSVLWASNHTKQYRHSFFADYYVQDV 116
DB      1939 SIITDYNESGLLQTAFLGTSRRVLFKYRRQTFELSEILDST-----RVSF---TYDETA 1990

QY      117 ESKSVPLMPDQEG---DIQYQMSVGNNTIAF-----VRENDLYVWDNGTTRITDD 166
DB      1991 GVLKTNVQLSD--GFICTIRYQIGPLIDRQIFRSEDGMVNAFYSYDNS--FRVTSM 2046

QY      167 GGPDMFHGVDPDMEIEELGDRVALWFSPDGEY-LAYLSFNE---TGVPYTYT----- 214
DB      2047 QGVINETPLDLYQFDDISGKVEQF-----GKFGVIYDINOIISTAVMTYTKHFDAGH 2102

QY      215 ---VQYMDNOEIAPAYPMELKIRYKPSQTNPTVTLSSLNIAKSEVKQAP----- 262
DB      2103 IKEIQY-----ELFRSLMYWITIYDYMGR-----VTKREIKIGFPANTTKYAY 2146

QY      263 ----IDAFESTDLIGEVAMLTDTHTTAAKAFNRVQDQKVAADVTASNKATVISD---R 316
DB      2147 EYDVEDQLOQTVYLNKIMWRNYDNLGNLHLPNSNSARLTPLRYDLDRITRLGQVQYR 2206

QY      317 DGTGHWL-----DNLISMKYIGIPKPSDKDAYVIDISDHSQWAHLVLPFVSGG 364
DB      2207 LDEDFGLRQRGTEIFBYSSKGLLTRVY-----SKSGGWTVIYRYDGLGR 2250
```

Qy 365 EPIPLTKGDWEVTSILSDQERQLVYVYLSLTOHHSTERHLSVSVSTFAVTPLVDDTVAAV 424
Db 2251 -----RVSSKTSILGQHLQF-FYADLTYPTRITHVY--NHSSEITSLSYLDLQGLH 2297
Qy 425 WSASFANSYGYLYTYGGPDVPYQELVYTNSTKPLRTITDNKVLQIKDYALPNITYFE 484
Db 2298 FAMEISSGDEFYIA-----SDNTGTPLAVFSSNGLMLKQIQTAYGEI-YFD 2343
Qy 485 LPLPSGETLNMORLPPGFS-----PDKKYPILFTPYGPGGAQEVTKR--WQ 529
Db 2344 -----SNIDFQLVIGFHGGLYDPLTKLIHFGERDYDILAGRWTPDI-EIWKRIKD 2394
Qy 530 ALNFKAYVASDSE-----LEYVT---WTV-----DNRGTGFKGRKFRSAVTRQLG 571
Db 2395 PAPENLYMFRNNPASKIHDVKDYITDVNSMLVTFGFHLHNAIPGFPVPKF-----D 2446
Qy 572 LLEAQDIYAAQAANIPIWDADHIGI-----VLEKDSGAFTLGYI-----TAPVSDWRFY 599
Db 2447 LTPESYELVKVSKQWDDIPPI-----FGVQQQVARQAKAFSLGKMAEVQVSRRRAGGAQSW 2502
Qy 600 GWSFGGYLTSK-----VLEKDSGAFTLGYI-----TAPVSDWRFY-----D 635
Db 2503 LW-----FATVKSLLGKGVMLAVSQGRVQTNVINTANEDCIKVAALVNNAFYLENLHFTIE 2558
Qy 636 SMYTERYMKLTSTNEEGYETSAVRKTDGFKNVEGGFLIQHGTGDDNVHFQNSAALV----- 691
Db 2559 GKDTHYFIKT--TTPE-S-DLGTLRLTSGRKALENGI-----NVTVSQSTTVVNGRT 2606
Qy 692 ----DLLMGDGVSPKLSHQWFTSDHGISVHGGVFLYKQARKLYQEKNRQTVLMHQ 747
Db 2607 RRFADVEMQFGAL--ALHVR-----GMTLDEEKARILEQAROR---ALARA 2648
Qy 748 WTK-----KOLEE 755
Db 2649 WAREQQRVRDGE 2661

RESULT 12
US-11-096-051-10
; Sequence 10, Application US/11096051
; Publication No. US20050244868A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacLachlan, Timothy K
; APPLICANT: Rastelli, Luca
; APPLICANT: Vernet, Corine
; APPLICANT: Etenberg, Seth
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use
; FILE REFERENCE: Attorney Docket No. Cura 967
; CURRENT APPLICATION NUMBER: US/11/096,051
; CURRENT FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: 10/038,854
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 10/455,772
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: 60/557,978
; PRIOR FILING DATE: 2004-03-30
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 10
; LENGTH: 2721
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-096-051-10
Query Match 3.1%; Score 123.5; DB 7; Length 2721;
Best Local Similarity 20.0%; Pred. No. 0.57;
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;

Qy 60 NIVT--NRSRTLI PADKPTGKEA-FNYWIHPDLSSVLWASNNHTKQYRHSFPADYVQDV 116
Db 1945 SIITDYNEEGLQLTAFLGTSRRVLFKYRQTRLSEILDST-----RVSP---TYDETA 1996

Qy 117 ESLSKSVPLMPDQEG---DIOYAQWSPVGNITAF-----VRENDLYVMNGTVTTRITDD 166
Db 1997 GVLKTVNLQSD--GFICTIRYQIGPLIDRQIFRSEDGMVNARFDYSYDNS---FRVTSM 2052
Qy 167 GGPDMFHGVPDWIYEEBILGDRVALWSPDGEV-LAYLSFNE---TCVPYTYT----- 214
Db 2053 QGVINETPLFDILYQFDDISGRKEQF-----GKGVYIYDINOIISTAVMTYTKHFDAGR 2108
Qy 215 ---VOYTMDOEIAPIAPYMWELKIRYKPVKSOTNPVTVLSLNNIASKEVKQAP----- 262
Db 2109 IKEIQY-----EIFRSLMWYITQYDNMGR-----VTKREIKIGPFANTTKYAY 2152
Qy 263 ---IDAFESTDLIIGEVAWLTDTHTTVAAKAFNRVQDQKVAVDTASNKATVSD---R 316
Db 2153 EYDVGOLQTVYLNKIMMYRYNDLNGNLHLNPSNSARLTPLRYDURDRITRGDVOYR 2212
Qy 317 DGTGDLW-----DNLLSMKYIGIPKPSDKDAYVIDISDHSGWAHLVLPVSGG 364
Db 2213 LDEGDFLRQGRTEIFEYSSKGLLTRVY-----SKSGWTVIYRYDGLGR 2256
Qy 365 EPIPLTKGDWEVTSILSDQERQLVYVYLSLTOHHSTERHLSVSVSTFAVTPLVDDTVAAV 424
Db 2257 -----RVSSKTSILGQHLQF-FYADLTYPTRITHVY--NHSSEITSLSYLDLQGLH 2303
Qy 425 WSASFANSYGYLYTYGGPDVPYQELVYTNSTKPLRTITDNKVLQIKDYALPNITYFE 484
Db 2304 FAMEISSGDEFYIA-----SDNTGTPLAVFSSNGLMLKQIQTAYGEI-YFD 2349
Qy 485 LPLPSGETLNMORLPPGFS-----PDKKYPILFTPYGPGGAQEVTKR--WQ 529
Db 2350 -----SNIDFQLVIGFHGGLYDPLTKLIHFGERDYDILAGRWTPDI-EIWKRIKD 2400
Qy 530 ALNFKAYVASDSE-----LEYVT---WTV-----DNRGTGFKGRKFRSAVTRQLG 571
Db 2401 PAPENLYMFRNNPASKIHDVKDYITDVNSMLVTFGFHLHNAIPGFPVPKF-----D 2452
Qy 572 LLEAQDIYAAQAANIPIWDADHIGI-----VLEKDSGAFTLGYI-----TAPVSDWRFY 599
Db 2453 LTPESYELVKVSKQWDDIPPI-----FGVQQQVARQAKAFSLGKMAEVQVSRRRAGGAQSW 2508
Qy 600 GWSFGGYLTSK-----VLEKDSGAFTLGYI-----TAPVSDWRFY-----D 635
Db 2509 LW-----FATVKSLLGKGVMLAVSQGRVQTNVINTANEDCIKVAALVNNAFYLENLHFTIE 2564
Qy 636 SMYTERYMKLTSTNEEGYETSAVRKTDGFKNVEGGFLIQHGTGDDNVHFQNSAALV----- 691
Db 2565 GKDTHYFIKT--TTPE-S-DLGTLRLTSGRKALENGI-----NVTVSQSTTVVNGRT 2612
Qy 692 ----DLLMGDGVSPKLSHQWFTSDHGISVHGGVFLYKQARKLYQEKNRQTVLMHQ 747
Db 2613 RRFADVEMQFGAL--ALHVR-----GMTLDEEKARILEQAROR---ALARA 2654
Qy 748 WTK-----KOLEE 755
Db 2655 WAREQQRVRDGE 2667

RESULT 13

US-11-096-051-8
; Sequence 8, Application US/11096051
; Publication No. US20050244868A1
; GENERAL INFORMATION:
; APPLICANT: Kekuda, Ramesh
; APPLICANT: MacLachlan, Timothy K
; APPLICANT: Rastelli, Luca
; APPLICANT: Vernet, Corine
; APPLICANT: Etenberg, Seth
; TITLE OF INVENTION: Ten-M3 Polypeptides and Polynucleotides and their Methods of Use
; FILE REFERENCE: Attorney Docket No. Cura 967
; CURRENT APPLICATION NUMBER: US/11/096,051
; CURRENT FILING DATE: 2005-03-30
; PRIOR APPLICATION NUMBER: 10/038,854


```
; PRIOR FILING DATE: 2001-12-31
; PRIOR APPLICATION NUMBER: 10/455,772
; PRIOR FILING DATE: 2003-06-04
; PRIOR APPLICATION NUMBER: 60/557,978
; PRIOR FILING DATE: 2004-03-30
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 8
; LENGTH: 2725
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-096-051-8

Query Match          3.1%; Score 123.5; DB 7; Length 2725;
Best Local Similarity 20.0%; Pred. No. 0.57;
Matches 171; Conservative 105; Mismatches 290; Indels 287; Gaps 45;

QY 60 NIWT--NESRTIIPADKIPTGKEA--FNYWIHPDLSSVLWASNHTKQYRHSFPFADYVQDV 116
Db 1949 SIITDYNREGLLLOTAFLGTSRRVLFKYRQTRLSLLYDST-----RVSP--TYDETA 2000

QY 117 ESKSVPLMPDQEG---DIQYAGSPVGNITAF-----VRENDLYVWDNGTTRITDD 166
Db 2001 GVLKTNLQSD--GFICTIRYRQIGPLIDRQIFRSEDCMVNARFDYSYDNS--FRVTSM 2056

QY 167 GQPMFHGVPDWIYEEIEILGDYALWFSFGY--LAYLSFNE---TGVPTYT----- 214
Db 2057 QGVINETPLIDLYQDDISGKVEQP-----GKFGVIYDINQIISTAVNTYTKHFDAGR 2112

QY 215 ---VOYMDNQIAPAYPWEIKIRYPKVQTNPTVTLNLINASKVEKQAP----- 262
Db 2113 IKEIQY-----EIFRSLMTWIIQYDNMGR-----VTKREIKIGPANTTKYAY 2156

QY 263 ----IDAFESTDLIIGEVAWLTDTHTTAAKAFNRVQDQKVAVDATASKATVISD--R 316
Db 2157 EYDVQDQLQTVVLYNEKIMRYNDLGNLHLLNPSNSARLTPLRYLDRITRLGDOVQR 2216

QY 317 DGTGDL-----DNLLSMKYIGPIKPSDKDAYIIDSDHSGWAHLVLPFVSGG 364
Db 2217 LDEDFGLRQGTETFEYSSKGLLTRVY-----SKSGMTVIYRIGLGR 2260

QY 365 EPIPLTKGDWEVTSILSDQEROLVYVYLSLQHHSTERHLYSVSYSTFAVTPLVDDTVAA 424
Db 2261 -----RVSGSKLQGHLP--FYADLTPTRIHYV--NHSSEITSLYDLOGLH 2307

QY 425 WSASFSGANSYYLITGGPDVPVQELYTNTSTKPLRTITDNNAKLEQIKDYALPNITYPE 484
Db 2308 FAMEISSGDEFYIA-----SDNTGTPLAVFSSNGMLMKQIQVTAYPEI--YFD 2353

QY 485 LPLPSGETLVNMORLPPGFS-----PDKKYPIPLFTPYGPGQAQEVTKR--WQ 529
Db 2354 -----SNIDFQLVIFGHGGLYDPLTKLIHFERGDYDILAGRWTTPDI--EIWKRI 2404

QY 530 ALNFKAYVADSE-----LEVVT-----WTV-----DNRGTGFKGRKFRSAVTRQLG 571
Db 2405 PAFNLYMERNPNPASKIHVDKDIITDVNSWLVTFGFHLHNAIPGPVPKF-----D 2456

QY 572 LLEAEQDIYAAQQAANIPWIDADHIGI-----FGVQQQVQARQAKAFLSLGKMAEVQV 599
Db 2457 LTESPELVKSKQWDDIPPI-----VLEKDSGAFTLGI-----TAPVSDWRFY----- 635

QY 600 GWSFGGYLTSK-----VLEKDSGAFTLGI-----TAPVSDWRFY----- 635
Db 2513 LW----FATVKSILIGKVMVAVSQGRVQTVNLVIANEDCIIKVAAVLNNAFYLENLHFT 2568

QY 636 SMYTERYKTLSTNEGYETSAVRKTDGFKNKGFGFLIQHGTGDDNVHVFQNSAALV---- 691
Db 2569 GKDTHYFIKT--TTPSS--DIGTLRLTSGRKALENGI-----NTVSQSTTVNGRT 2616

QY 692 ----DLIMGDGVSPEKLHSQWFTSDSHGISYHGGGVFLYKQLARKLYQEKROTQVLMHQ 747
Db 2617 RRPADEVMOQFAL--ALHVRV-----GMTLDEKARIIEQARQ--ALARA 2658
```

```
QY 748 WTK-----KLEE 755
Db 2659 WAREQORVDRGEE 2671

RESULT 14
US-10-995-561-625
; Sequence 625, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 625
; LENGTH: 1259
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-625

Query Match          3.0%; Score 122.5; DB 6; Length 1259;
Best Local Similarity 21.3%; Pred. No. 0.23;
Matches 135; Conservative 75; Mismatches 205; Indels 219; Gaps 34;

QY 181 EBEILGDYALWLF--SPDGEYLAVLSFNETGPTVTVTVQYYMDNQE-----IAPAYPW 230
Db 297 EEWVHADQSSCTFDNLSPGLE-----NVSVYTVK---DDKESVPISDTIIPAVPP 344

QY 231 ELKIRYPKVQTNPTVTL-----LLNIASKEVQKAPIDAFESTDLIIGEVAWLTDTHTTVA 287
Db 345 PDLRFTNIGPDMRTVWAPPSSIDLTLNPLRISPVKNEE-----DVAELISISDANA 397

QY 288 AKAFNRVQDQKVAVDTA--SNKATVISDRDGTGWLNDLLSMKYIGIPKPSDKDAYI 345
Db 398 VVLTNLLPCTEYVVSVSVEVQEHSTPLRQKT-----GLDSPTGID--FS 442

QY 346 DISHSGWAHLVFP---VSG-----GEP-----IP-----LTGQWEVT 377
Db 443 DITANSFTVH--WIAPRATITGYRIHHPHFSGRPREDRVPHSRNSITLTLNTPGTEYV 501

QY 378 SILSDQERQVYVLSLQHHSTERHLYSVSYSTFAVTPLVDDTVAA-----YWSASFS 430
Db 502 SIVALNGREESPLLIGQQ-----STVSDVPRDLEVVAAATPSSLISWDAP--A 547

QY 431 ANSGYVILTYG--GPDVPYQELYTNTSTKPLRTITDNNAKLEQIKDYALP----- 478
Db 548 VTVRYRYIYGETGNSPVQE--FTVPGSKSTATISG---LKPQVDYITTVYAVTGRGDS 602

QY 479 -----NTTY--FELPLPSGETLVNQ-----RLPPGSPDKKYPIPLFTPYGPGQAQEV 524
Db 603 PASSKEPISINRYTEIDKPSQMOVTDQDNSISVKMLPSSSPVTVGYRTVTPKNGPPTKT 662

QY 525 TKRQALNFKAYVASDESELEYVTVVNDNRGTGFKGRKFRSAVT---RQLGLLEAEQIYA 581
Db 663 KTAGPDQTEMTIEGLQPTVEYVVSVAQNPSGESQPLVQVAVTNIDRPNKGLAFTD----- 717

QY 582 AQAANIPIWDADHIGIWSFGYLTSKLEKDSGAFTLVITAPVSDWRFRYDSMYTER 641
Db 718 -----VDVDSIKI--AW-----ESPQG-----QVSR 736

QY 642 YMKLTSTNEEG-----YETSAVRKTDGFKNKGFGFLIQHGTGDDNVHVFQNSAALV---- 676
Db 737 YRVYSSPEDGIELHFPAPDGEEDTABLQRLPGSEYTVSVWALHD---DMESQPLI--G 791

QY 677 TGDNDVHFQNSA--ALVDLLMGDGVSPKLSQW 708
Db 792 T-----OSTAIPAPTDLKFTQ--VTPTSLSAQW 817
```

```
RESULT 15
US-10-995-561-628
; Sequence 628, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 628
; LENGTH: 1286
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-995-561-628

Query Match      3.0%; Score 122.5; DB 6; Length 1286;
Best Local Similarity 21.3%; Pred. No. 0.23; Mismatches 219; Gaps 34;
Matches 135; Conservative 75; Indels 219; Gaps 34;

Qy 181 EEEILGDRYALWF---SPDGEYLAYLSFNETGVPTTYVQYMDNQE-----IAPAYPW 230
Db 297 EEVHADQSSCTFDNLSGLEIY-----NVSYYTVK---DDKESVPISDTIIPAVPP 344

Qy 231 ELKIRYKVSQTNPTVTLUS-----LNIAKVKQAPIDAFESTDLIIGEVAMLTHTHTYA 287
Db 345 PTDLRFTNIGPDTWRVTWAPPPSIDLTNLFVRYSPVKNEE-----DVAELSIKSPSDNA 397

Qy 288 AKAFNRVQDQKVAVDTA--SNKATVISDRDGTGDLNLLSMKYIGPIKPSDKDAYI 345
Db 398 VVLNLLPGTEYVSVSVYQCHESPTLRGQKT-----GLDSPTGID--FS 442

Qy 346 DISHSGWAHLYLFP---VSG-----GEP---IP-----LTKGDMVEVT 377
Db 443 DITANSFTVH-WIAPRAITGVRIRHHPHFSGRPREDRVPHSRNSITLTNLTGTEYV 501

Qy 378 SILSIDQERQLVYVYLSLTHSTHERLYSVSYSTFAVTLVDDTVAA-----YWSASF 430
Db 502 SIVALNGREESPLLTGQQ-----STVSDVPRDLEVAATPTSLLSMDAP-A 547

Qy 431 ANSGYYILTYG--GPDVPYQELTYTNTSKPLRTITDNAKVLEQIKDYALP----- 478
Db 548 VTVRYRITYGETGNSPVQE--FTVPGSKSTATISG----LXPGVDYITVYAVTGRGDS 602

Qy 479 -----NITY-FELPLPSGETLVNQ-----RLPPGFSPPKKYPILFTPYGPGQAEV 524
Db 603 PASSKPIISINYRTEIDKFSQMQVTDVQNSISVKWLPSSPVTVGYRVTTTPKNGPGPTKT 662

Qy 525 TKRQALNFKAVASDSLELYVTWTDNRGCTGKGRKFRSAVT---ROIGLLEAEDQIYA 581
Db 663 KTAGPDQTEMTIEGLQPTVEYVSVYAQNPSGESQPLVQTAVTNIDRPKGLAFTD----- 717

Qy 582 AQQAANIPIWDADHIGIWCWFGGYLTSKVLEKDSGAFTLGVITAPVSDWRFPYDSMYTER 641
Db 718 -----VDVDSIKI-AW-----ESPOG-----QVSR 736

Qy 642 YMKTLTSTNEEG-----YETSAVRKTDGFKNVGEGFLIQHG 676
Db 737 YRVYSSPEDGIHELFPAPDGEEDTAELQGLRPGSEYTVSVVALHD---DMESQPLI--G 791

Qy 677 TGDDNVHFQNSA--ALVDLLMGDGVSPKLSQW 708
Db 792 T-----QSTAIPAPTDLKFTQ-VTPTSLSAQW 817
```